Discussion Papers No. 291, December 2000
Statistics Norway, Research Department

Arvid Raknerud and Rolf Golombek

Exit Dynamics with Rational Expectations

Abstract:
We develop an econometric model for firm exit, using stochastic dynamic programming (SDP) as a starting point. According to SDP, the value of an operating firm can be written as the sum of (i) the net present value of continuing production if the firm is committed to a future exit date, and (ii) the value of the exit option. By approximating the option value by a simple function of its determinants, we derive an expression for the distribution of firm exit probabilities. The model is estimated by pseudo likelihood methods using panel data from the Norwegian Manufacturing Statistics. The applicability of the model is illustrated by assessing to what extent quotas on emissions of carbondioxide increase exits in manufacturing sectors.

Keywords: Exit dynamics, stochastic dynamic programming, option value, pseudo likelihood, dynamic panel data, random effects, environmental taxes.

JEL classification: C33, C51, C61, D21, Q38

Acknowledgement: We thank John Dagsvik, Søren Johansen, Tor Jakob Klette, and seminar participants at Statistics Norway, and University of Oslo for many useful comments. Financial support from the LOS Programme of the Norwegian Research Council is gratefully acknowledged.

Address: Arvid Raknerud, Statistics Norway, Research Department, P.O. Box 8131 Dep., N- 0033 Oslo, Norway E-mail: arvid.raknerud @ssb.no

Rolf Golombek, The Frisch Centre, University of Oslo, Gaustadalleen 21, N-0349 Oslo, Norway. E-mail: rolf.golombek@frisch.uio.no.

Discussion Papers comprise research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborate than a standard journal article by including intermediate calculation and background material etc.

Abstracts with downloadable PDF files of
Discussion Papers are available on the Internet: http://www.ssb.no

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
N-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1 Introduction

What causes a firm to exit? The standard answer from economists is low profitability. According to Marshall, "production is likely to come to a sharp stop ... [when] ... the price falls so low that it does not pay for the out of pocket expenses" (Marshall (1966), p. 349). While the Marshallian exit rule is myopic, modern economic theory assumes that the exit decision is based on both present and expected future revenues and costs. Moreover, firms have rational expectations. In its most refined version, the exit rule of a firm is derived from stochastic dynamic programming (SDP): The firm will stay operative as long as the expected present value of continuing production exceeds the value of closing down. This exit rule is derived under the assumption that the firm always chooses optimal actions, given the available information at each point in time.

The present paper develops an econometric model for firm exit, using SDP as a starting point. The model is estimated, and then applied to assess the degree to which new environmental taxes – through shifts in the profitability of firms – may cause more exits.

In Das (1992), an econometric implementation of a discrete choice SDP model of firm behavior is presented. Das considers the problem of whether to operate, hold idle, or close down production units in the cement industry. By following the specifications in the pioneering article by Rust (1987) on bus engine replacement, Das derives explicit choice probabilities as functions of primitive parameters. However, solution methods for discrete choice SDP problems with continuous state variables require approximating the problems with finite ones, which leads to high-dimensional state spaces. Although some attempts have recently been made to break this "curse of dimensionality" by means of randomization methods (see Rust (1997)), computational complexity remains a fundamental obstacle to the practical usefulness of SDP in econometrics. While SDP provides a general framework for interpreting and analyzing intertemporal decision problems, implementation of SDP models often require too severe limitations on the number of variables that can be included in the analysis. Moreover, the econometrician often has to assume representative agents with homogenous information sets.

Due to the obstacles in applying SDP, the literature offers alternative approaches. A well-known approach is that of Stock and Wise (1990), who analyze the retirement decisions of older employees. Stock and Wise suggest a sub-optimal decision criterion that is meant to approximate the decision rule of SDP. Within our setting, the Stock and Wise rule implies that a firm finds, at t, the time to exit that maximizes expected profits, given the current information of the firm. The firm then "commits" to this time of exiting, say, $t + e_t^*$.

There is a fundamental problem with the Stock and Wise approach: A firm may obtain new information between t and $t + e_t^*$ that affects the desirability of exiting. While a firm is free to adjust the time to exit as long as it remains active, it gives up the possibility to wait for new information when it exits. This lost option (exit is an absorbing state) has an opportunity cost that should be included as part of the cost of exiting. Hence, the value of continuing production can be written as a sum of two terms: (i) the present value of exiting at $t + e_t^*$, and (ii) the option value – the value of keeping the exit option alive.[1]

While the sum of (i) and (ii) equals the value of continuing production according to SDP, the Stock and Wise approach captures (i) only. In our econometric model, we extend the Stock and Wise approach by approximating the value of the exit option by a simple function of the state variables. This approach is in line with Pesaran and Smith (1995), who propose to integrate dynamic optimization theory and empirical analysis by approximating Lagrange multipliers (derived from the stochastic maximum principle) by simple functions of their determinants.

In order to analyze the dynamic programming problem, we need a model for firm profitability. Firstly, we split profits into a revenue component and four cost components: capital, labor, energy and material costs. Secondly, each component is factorized into price and quantity (physical amount). Conditional on prices, log-output and log-factor use are assumed to evolve according to a VAR process. The VAR model implies a particular "history dependence": the firm's choice of factor inputs depends on current prices as well as lagged inputs and output. We show that the VAR model can be interpreted as an approximation to a reduced form relation derived from a structural model with exogenous prices and adjustment costs.

An important computational advantage of our approach is to allow for a high dimensional state vector. We also incorporate systematic heterogeneity between firms by including a firm-specific productivity parameter. The firm-specific productivity parameter is known to the firm, but not to the econometrician.[2] This causes a self-selection problem: both the firm's profit realizations and exit decisions depend on a common latent variable.[3]

The econometric model is estimated by pseudo likelihood methods, using panel data from the Norwegian Manufacturing Statistics for three sectors. In general, the data are explained well by

[1] We use the term "option value" similarly as in the finance literature, where a call option denotes a right, but not an obligation, to buy an asset at some future time. On the other hand, Stock and Wise (misleadingly) use the term "option value" to denote *(i)*. See also Dixit and Pindyck (1994), who decompose the value of a firm's investment opportunity into the net present value and the option value.

[2] For a more advanced approach of modelling firm heterogeneity, see Pakes and Ericson (1998), who apply (in a different setting) two models of heterogeneity: one with passive Bayesian learning in which firms are endowed (at birth) with an unknown value of a productivity parameter (observed profit realizations contain information on the value of the parameter), and one with active exploration in which the firm knows the current value of the productivity parameter, but the value of the parameter depends on stochastic outcomes of the firm's investments.

[3] Olley and Pakes (1996) analyze a similar selection problem – the relationship between unobserved productivity and the shutdown decision – in order to estimate the parameters of a production function for the US telecommunications equipment industry.

the model. The applicability of the model is then illustrated by assessing to what extent quotas on carbondioxide (CO_2) emissions cause increased exits.

Our main results are *(i)* the model improves goodness-of-fit significantly compared with a model with a purely myopic exit rule (no option value); *(ii)* there is, however, no conclusive evidence that joint (pseudo likelihood) estimation of the state process and the exit probabilities significantly improves predictions of exit compared with separate estimation (in one of the three sectors, McFadden's ρ^2 increased by 50 per cent with joint estimation, whereas ρ^2 remained virtually unchanged in the two other sectors); *(iii)* the individual random effect (the productivity parameter) has an impact on firm-specific exit probabilities primarily through profitability, not through the option value; *(iv)* in two sectors, we find strong evidence of a firm-specific common trend, i.e. a stochastic trend not caused by trends in industry-wide prices; and (v) with a quota price of USD 25 per tonne of CO_2 – a rough estimate of the price of CO_2 quotas following from the Kyoto agreement – aggregate exit probabilities increase only slightly.

The rest of the paper is organized as follows: In Section 2, the firm profitability model is presented. The exit decision problem of the firm is motivated and modeled in Section 3. Section 4 contains empirical applications: we first present the results of the estimation and then use the estimated model to examine the impact on firm exit following from imposing quotas on CO_2 emissions. Section 5 concludes the paper. The pseudo likelihood function and a detailed outline of the estimation algorithm are presented in Appendix A.

2 Firm profitability

In this section, we present the stochastic process which generates the profitability of the firms.[4] Let π_t^i denote profit of firm i in year t. We assume that π_t^i can be decomposed into five revenue and cost components, which we collect in the vector $y_t^i = (y_{tQ}^i, y_{tK}^i, y_{tL}^i, y_{tE}^i, y_{tM}^i)'$, where y_{tQ}^i denotes total revenue and y_{tK}^i, y_{tL}^i, y_{tE}^i, and y_{tM}^i denote capital, labor, energy, and material costs, respectively. The components of y_t^i are measured in real prices, using the consumer price index as a common deflator.

Profit can be written as

$$\pi_t^i = \iota' y_t^i, \tag{1}$$

[4] The term "firm" is used in the sense of plant or establishment.

with $\iota = (1, -1, -1, -1, -1)'$. Furthermore

$$\begin{aligned} y^i_{tQ} &= P_{tQ} X^i_{tQ} \\ y^i_{tj} &= P_{tj} X^i_{tj} \qquad j = K, L, E, M, \end{aligned} \tag{2}$$

where P_{tQ} is the industry-wide real price of output, X^i_{tQ} is output from firm i, P_{tj} is the industry-wide real price of factor input j, and X^i_{tj} is the corresponding factor use of firm i. Let

$$\begin{aligned} x^i_t &= (\ln X^i_{tQ}, ..., \ln X^i_{tM} \\ p_t &= (\ln P_{tQ}, ..., \ln P_{tM})'. \end{aligned} \tag{3}$$

We assume that $(x^{i\,\prime}_t, p_t{}')'$ is a vector autoregressive process, with p_t being an exogenous random walk:

$$\begin{aligned} x^i_t &= \mu + \Phi x^i_{t-1} + \Gamma p_t + \Upsilon v^i + \varepsilon^i_t \\ p_t &= p_{t-1} + \eta_t \end{aligned} \tag{4}$$

where μ is a 5×1 intercept vector, Φ and Γ are 5×5 matrices of autoregressive coefficients, v^i is a scalar random effect – known by the firm but not by the econometrician, Υ is a 5×1 factor loading matrix[5], and ε^i_t and η_t are Gaussian white noise vectors:

$$\begin{aligned} \varepsilon^i_t &\sim \mathcal{IN}(0, \Sigma) \\ \eta_t &\sim \mathcal{IN}(0, Q). \end{aligned} \tag{5}$$

Since the covariance matrix Q can be estimated directly from time-series data on prices, henceforth Q will be considered as a known parameter.

To complete the statistical specification of (4), we need to specify the joint distribution of v^i and $x^i_{s_i}$, where s_i is the initial observation year of firm i. Because firms are self-selected into the sample (through survival), this distribution depends on both the entry and the exit process. However, it is shown in Appendix A, Proposition 1, that the joint distribution of v^i and $x^i_{s_i}$ affects the estimation only through its contribution to the distribution of v^i given *all* the time-series data $x^i_{s_i}, ..., x^i_{T_i}$, where T_i is the last year firm i is observed. Under very general conditions, when $T_i - s_i$ is moderate or large, this distribution will be approximately normal and independent of the *cross-sectional* distribution of v^i given $x^i_{s_i}$ (see Lehmann (1983), Ch. 7). This suggests that the multivariate normal distribution may be a useful approximation:

$$v^i | x^i_{s_i} \sim \mathcal{IN}(\psi'(x^i_{s_i} - E\{x^i_{s_i}\}), 1), \tag{6}$$

[5] See, for example, Anderson (1984), Ch. 14, for an introduction to factor analysis.

where ψ is a 5×1 vector of regression coefficients and $E\{x_{s_i}^i\}$ is the unconditional expectation, which can be estimated directly from the cross-section of initial observations. Since v^i enters (4) through the term Υv^i, the standardization of the variance in (6) entails no loss of generality. Similarly, the restriction $E\{v^i\} = 0$ is necessary to identify the intercept μ.

The assumption that real output and input prices follow a random walk is obviously an over-simplification. These variables show trending behavior, but the trends are not linear, or even monotone. For example, in all industries considered there are downward trends in real wages from 1976 to the early 1980s, followed by upward trends. These patterns suggest stochastic trends. Thus, a simple I(1) model may provide a reasonable approximation.

To motivate the equations (4), it may be useful to try to relate them to a more structural model. Assume that firms have CES-technology:

$$X_{tQ}^i = \left[\sum_{j \in \{K,L,E,M\}} \left(A_j^i X_{tj}^i \right)^{\nu} \right]^{\kappa/\nu} \qquad (\nu, \kappa) \in (0, 1),$$

where A_j^i is the productivity of factor j in firm i. If we allow for firm-specific productivity differences, but restrict these to be time-invariant and (Hicks-) neutral (affecting all inputs in the same way):

$$A_j^i = c_j V^i,$$

profit maximization conditional on given prices leads to the factor demand functions (on log-form)

$$\begin{aligned} x_{tj}^i &= \frac{1}{1-\kappa} \ln \kappa + \frac{\rho c_j}{1-\nu} + \frac{1}{1-\kappa} \ln P_{tQ} + \frac{1}{\nu - 1} \ln P_{tj} + \frac{\kappa}{1-\kappa} \ln V^i \\ &\quad + \frac{\nu - \kappa}{\nu(\kappa - 1)} \ln \left[\sum_{j \in \{K,L,E,M\}} \exp\left(\frac{\nu}{\nu - 1} \ln(P_{tj}/c_j) \right) \right], \qquad j = K, L, E, M, \end{aligned}$$

and the supply function (on log-form)

$$\begin{aligned} x_{tQ}^i &= \frac{\kappa}{1-\kappa} \ln \kappa + \frac{\kappa}{1-\kappa} \ln P_{tQ} + \frac{\kappa(\nu - \kappa)}{(1-\nu)(\kappa - 1)} \ln V^i + \\ &\quad \frac{\nu - \kappa}{\nu(\kappa - 1)} \ln \left[\sum_{j \in \{K,L,E,M\}} \exp\left(\frac{\nu}{\nu - 1} \ln(P_{tj}/c) \right) \right]. \end{aligned}$$

These first-order (equilibrium) conditions are moderately non-linear in p_t, and can be linearized to give

$$\widetilde{\beta}_1' x_t^i + \widetilde{\beta}_2' p_t + \widetilde{\mu} + \widetilde{\Upsilon} v^i \approx 0 \tag{7}$$

for suitably chosen 5×5 matrices $\widetilde{\beta}_1$ and $\widetilde{\beta}_2$, 5×1 vectors $\widetilde{\mu}$ and $\widetilde{\Upsilon}$, and with $v^i = \ln V^i$. If the adjustment toward equilibrium is driven by a first-order error correction:

$$\Delta x_t^i = \widetilde{\alpha}(\widetilde{\beta}_1' x_{t-1}^i + \widetilde{\beta}_2' p_{t-1} + \widetilde{\mu} + \widetilde{\Upsilon} v^i) + \widetilde{\varepsilon}_t^i, \tag{8}$$

we obtain, through reparametrization, a model of the form (4).[6] The interpretation is that, while (7) defines the underlying equilibrium relations, firms being out of equilibrium adjust according to the adjustment mechanism (8) to bring the variables back to equilibrium.

Note that the parameters of the structural model cannot be identified from (4) because the parameters in the latter model depend on the adjustment matrix $\widetilde{\alpha}$; e.g. $\Phi = I + \widetilde{\alpha}\widetilde{\beta}_1{}'$. However, the model (4) should not be regarded as a tool for estimating production functions, but as a mean to facilitate predictions on future profitability. Furthermore, to the extent that changes in prices are correlated with industry-wide technological change, our model will capture this effect through Γ. Thus, one should be careful in interpreting Γ.

Model (4) is valid under more general conditions than (8), which states that there are exactly five cointegration (equilibrium) relations, that is, $\widetilde{\beta}_1' x_{t-1}^i + \widetilde{\beta}_2' p_{t-1}$. For example, one strand of the literature on firm dynamics claims, in accordance with Gibrat's law, that idiosyncratic productivity growth is independent of firm size.[7] According to one interpretation of Gibrat's law, cumulative innovations arising from R&D investments lead to stochastic trends in the state variables on log-scale (see e.g. Klette and Griliches (2000)). In our setting, this means that the number of cointegrating relations (i.e. the rank of $\Phi - I$) is less than five. We will return to this issue in Section 4.

3 The exit decision

For any random vector X_s^i, we will use the notation $E_t\{X_s^i\}$ to denote the conditional expectation of X_s^i with respect to the information set I_t^i of firm i at time t:

$$I_t^i = \{v^i, p_{\widetilde{s}_i}, x_{\widetilde{s}_i}^i, ..., p_{t\wedge\widetilde{T}_i}, x_{t\wedge\widetilde{T}_i}^i\},$$

where $\widetilde{s}_i$ is the birth date and $\widetilde{T}_i$ is the exit date of firm i.[8] For any t in which the firm is operating, I_t^i can be described as knowledge about random variables which have been realized up to year t. Informally, I_t^i is the subjective information of firm i "at the end" of year t. On the other hand, the "objective" information set of the econometrician, I_t^o, consists of random variables $\{p_0, .., p_T\}$ and $\{x_s^i\}$ for $s_i \leq s \leq T_i$ and $i = 1, ..., N$, where $s_i = \widetilde{s}_i \vee 0$ and $T_i = \widetilde{T}_i \wedge T$. That is, the econometrician's observation period is $[0, T]$. In particular, he or she does not observe v^i.

[6] This is an instance of the celebrated representation theorem of Engle and Granger (1987).
[7] See Sutton (1997) for an overview of this literature.
[8] We use the standard notation $s \wedge t \equiv \min(s,t)$ and $s \vee t \equiv \max(s,t)$.

At the end of year t, the firm decides whether to close down or not. If the firm has a *finite* planning horizon $[t, t+H]$, this is a finite horizon optimal stopping problem. In order to analyze this decision problem, we first, following Stock and Wise (1990), look at a much simpler problem: Assume that the firm, given that it chooses not to exit, commits to a future exit time. If the firm chooses to continue production until $t+e$ $(e>0)$ and then closes down, it obtains the cash flow $\{\pi^i_{t+1}, .., \pi^i_{t+e}\}$ from production and, upon exiting, the scrap value S_{t+e}. If $V^i_t(e)$ is the total value of this exit-strategy, then

$$V^i_t(e) = \Pi^i_t(e) + \delta^e E_t\{S^i_{t+e}\}, \tag{9}$$

where

$$\Pi^i_t(e) = \sum_{k=1}^{e} \delta^k E_t\{\pi^i_{t+k}\},$$

δ is the discount factor, and $E_t\{S^i_{t+e}\}$ is the expected scrap value. Note that, in our setting, current profit π^i_t is immaterial for the firm's decision because it is earned regardless of whether or not the firm exits at the end of year t.

The firm can calculate the net present value $\Pi^i_t(e)$ conditional on its information set I^i_t using (4) as a prediction model. In particular, from (1) we get

$$\Pi^i_t(e) = \iota'(\sum_{k=1}^{e} \delta^k E_t\{y^i_{t+k}\}). \tag{10}$$

Equations (2)-(3) and a well-known property of the log-normal distribution yield

$$E_t\{y^i_{t+k,j}\} = \exp\left(E_t\{x^i_{t+k,j} + p_{t+k,j}\} + \tfrac{1}{2}Var_t\{x^i_{t+k,j} + p_{t+k,j}\}\right), \qquad j = Q, K, L, E, M, \tag{11}$$

where $Var_t\{\cdot\}$ denotes the conditional variance relative to I^i_t. The moments in (11) can be obtained from $E_t\{(x^i_{t+k}{}', p_{t+k}{}')'\}$ and $Var_t\{(x^i_{t+k}{}', p_{t+k}{}')'\}$, which, in turn, can be calculated recursively from (4) conditional on v^i, x^i_t and p_t.

Because the scrap value S^i_t is unobserved by the econometrician, we have to make some simplifying assumptions. First, we assume that S^i_t can be decomposed into assets A^i_t yielding a random rate of interest $\widehat{\rho}^i_t$, and a random residual ζ^i_t:

$$S^i_t = A^i_t + \zeta^i_t, \tag{12}$$

where

$$A^i_{t+1} = (1 + \widehat{\rho}^i_{t+1})A^i_t.$$

A^i_t could be the value of some asset which can be disposed of upon exit, while ζ^i_t is the net revenue from all other revenues and costs incurred by exit (e.g. revenues from sale of production equipment, clean-up costs, compensation to employees). Our basic hypothesis is that temporal variations in ζ_t^i and $\widehat{\rho}_t^i$ have a purely transient character. In particular, for all $e > 0$, and conditional on I_t^i,

$$E_t\{\psi_{t+e}^i\} = 0 \tag{13}$$

$$E_t\{\widehat{\rho}_{t+e}^i\} = \rho. \tag{14}$$

If $\widehat{\rho}_{t+e}^i$ and $\widehat{\rho}_{t+k}^i$ are conditionally independent for $e \neq k$ $(e, k > 0)$, equations (13)-(14) yield, with $\delta = \frac{1}{1+\rho}$,

$$\delta^e E_t\{S_{t+e}^i\} = \begin{cases} A_t^i & e > 0 \\ A_t^i + \zeta_t^i & e = 0. \end{cases} \tag{15}$$

The assumptions leading to (15) are restrictive, but warranted in the absence of direct data on scrap value.

Following Stock and Wise, we define

$$V_t^{i*} = V_t^i(e_t^*), \tag{16}$$

where

$$e_t^* = \underset{e \in \{1,\dots,H\}}{\arg\max}\ V_t^i(e). \tag{17}$$

Thus, given that the firm at the end of year t commits to a future exit time, V_t^{i*} is the value of remaining operative when the firm chooses this exit time so as to maximize expected discounted profits.[9]

The decision rule proposed by Stock and Wise seeks to approximate the correct value of remaining active, which can be derived using SDP. According to their (sub-optimal) rule, the firm will close down if $S_t^i > V_t^{i*}$. However, this is a biased – and typically poor – approximation.[10] In particular, the criterion ignores the option value: the decision to postpone exit means that the firm has an opportunity to exit at $t + e_t^*$, but it is not obliged to do so. This option has a value to the firm, λ_t^i, given by

$$\begin{aligned} \lambda_t^i &\equiv E_t\{\max_{1 \le \widetilde{e} \le H} \sum_{k=1}^{\widetilde{e}} \delta^k \pi_{t+k}^i + \delta^{\widetilde{e}} S_{t+\widetilde{e}}^i\} - E_t\{\sum_{k=1}^{e_t^*} \delta^k \pi_{t+k}^i + \delta^{e_t^*} S_{t+e}\} \\ &\equiv W_t^i - V_t^{i*}, \end{aligned}$$

where $\widetilde{e}$ is a stopping time relative to the firm's information set I_t^i. The total value of remaining operative, W_t^i, can then be written as

$$W_t^i = V_t^{i*} + \lambda_t^i.$$

[9] Note that the infinite horizon case ($H = \infty$) can be covered provided $V_t^i(e_t^*)$ converges as $H \to \infty$. For example, with a 7 percent rate of interest and parameter values equal to any of the estimates reported in Section 4, convergence is achieved. This is analogous to the method of successive approximations for solving infinite horizon SDP problems; one then solves a sequence of finite horizon problems until the value function converges (see Rust (1994)).

[10] For a study of its mathematical accuracy in some relevant situations, see Stern (1997).

Note that $\lambda_t^i \geq 0$. In a deterministic model, $\lambda_t^i = 0$, since the firm obtains no new information as time passes.

Empirical evidence suggest that non-profitable firms may stay operative for a long time. This behavior can hardly be explained by the decision rule "close down if $S_t^i > V_t^{i*}$". However, it could be rationalized when the option value is taken into account. Evidence on firms' investment behavior also supports this view. For example, Dixit and Pindyck (1994), citing Summers (1987), claim that investment projects typically have an expected yield which is three or four times the cost of capital.

As discussed thoroughly in Dixit and Pindyck (1994), the larger the variability in profit, the higher the option value. The reason is that the option introduces an asymmetry which causes the expected gain of a positive shock to be larger than the expected loss of a negative shock. To see this in our case, first consider a positive shock to profit. A positive shock increases π_t^i, but also expected future profit, and hence V_t^{i*}. On the other hand, a negative shock decreases V_t^* accordingly. Because the firm can choose to exit at the end of t, future losses can be avoided. If positive and negative shocks occur with equal probability, an increase in the variability of profit increases the difference between the upside potential and the downside risk, and thus increases the option value.

Following the general approach of Pesaran and Smith (1995), we propose to approximate the option value λ_t^i by a function, $\widehat{\lambda}_t^i$. In general, because the state process is Markovian, λ_t^i depends on p_t, x_t^i, and v^i. However, below we will seek to reduce the dimension of the state space over which we approximate λ_t^i.

The above discussion showed that variability in y_t^i is the source of the option value. Because of log-normality, the variance in the components of y_t^i is proportional to the expected value of y_t^i. Hence, variation in the *levels* of the y_t^i components across i and t also reflects variation in the *variance* of profits, and thus in the variance of option values. Let $Z_t^i = (Z_{t1}^i, .., Z_{td}^i)'$ denote the first d principal components of y_t^i, that is, the orthonormal basis for the d-dimensional subspace with most variation in the y_t^i-data.[11] We propose the following linear approximation:

$$\lambda_t^i = \widehat{\lambda}_t^i + \xi_t^i, \tag{18}$$

where

$$\widehat{\lambda}_t^i = \beta_0^* + \beta^{*\prime} Z_t^i + \gamma^* v^i, \tag{19}$$

and ξ_t^i is the error of approximation (note that the error term ξ_t^i has a degenerate distribution conditional on I_t^i). We restrict the principal components to be uncorrelated with π_t^i (hence $d \leq 4$). This

[11] See Anderson (1984), Ch. 11, for an introduction to principal components analysis.

is necessary to avoid identification and interpretation problems: to the extent that profitability and option value are correlated, it is impossible to separate them solely on the basis of data.

The decision rule derived from SDP is: *Close down production if and only if*

$$S_t^i > W_t^i \Longleftrightarrow -(\widehat{\lambda}_t^i + \Pi_t^{i*}) > \chi_t^i, \tag{20}$$

where $\Pi_t^{i*} = \Pi_t^i(e_t^*)$ and χ_t^i is the difference between the error terms in (12) and (18).

Let z_t^i be the indicator that firm i closes down:

$$z_t^i = \begin{cases} 1 & \text{if } i \text{ closes down at } t \\ 0 & \text{if } i \text{ does not close down at } t. \end{cases}$$

If $F_t^i(\cdot)$ denotes the cumulative distribution function of the error term χ_t^i conditional on I_t^i, we get the following relationship:

$$P(z_t^i = 1|v^i, x_t^i, p_t, z_{t-1}^i = 0) = F_t^i(-\widehat{\lambda}_t^i - \Pi_t^{i*}). \tag{21}$$

Note that the present value Π_t^{i*} depends on the random effect v^i and the unknown time-series parameters $\theta = (\Phi, \Gamma, \Sigma, \mu, \Upsilon)$. To stress this dependence, we will sometimes use the notation $\Pi_t^{i*}(v^i; \theta)$.

Since the probability distribution $F_t^i(\cdot)$ is unknown, we will rely on pseudo likelihood methods (see Gourieroux and Montfort (1995), Ch. 8.4). In particular, we approximate $F_t^i(\cdot)$ by a logistic distribution with scale factor σ. The approximate exit probability then has the familiar and convenient form

$$\widehat{P}_\omega(z_t^i = 1|v^i, x_t^i, p_t, z_{t-1}^i = 0\,; \theta) = \frac{1}{1 + \exp\left\{-\left(\beta_0 + \beta' Z_t^i + \gamma v^i + \alpha \Pi_t^{i*}(v^i; \theta)\right)\right\}}, \tag{22}$$

where $\beta_0 = -\beta_0^*/\sigma$, $\beta = -\beta^*/\sigma$, $\gamma = -\gamma^*/\sigma$, $\alpha = -1/\sigma$, and $\omega = (\beta_0, \beta, \gamma, \alpha)$. In the logistic distribution (22), ω should be interpreted as a *pseudo true* vector-parameter: the value of the parameter that minimizes the Kullback discrepancy between the true distribution $P(\cdot)$ and the postulated class of distributions $\widehat{P}_\omega(\cdot)$. An estimator that converges in probability to the pseudo true ω is obtained by maximizing the pseudo likelihood jointly with respect to θ, ψ (the parameters of the initial distribution of v^i) and ω, that is, by maximizing the function obtained by replacing $P(\cdot)$ with $\widehat{P}_\omega(\cdot)$ in the true likelihood function.[12]

The main purpose of statistical inference is to estimate exit probabilities as functions of *observed* variables. This means that the random effect v^i must be integrated out of (22) with respect to its conditional distribution, given the data. Because the derivations are tedious, the joint pseudo likelihood function, as well as a detailed outline of the estimation algorithm, is presented in Appendix A.

[12] The pseudo likelihood approach is in line with Pesaran and Smith (1995) – although they do not explicitly interpret their inference problems in terms of pseudo likelihood and pseudo true values.

4 Applications

In this section, we first estimate the model and then simulate effects on firm exit following from the imposition of quotas on emissions of CO_2. We rely on raw data from Statistics Norway's Manufacturing Statistics, which provide annual observations on gross value of production, intermediates, wage costs, energy costs, etc. for all Norwegian manufacturing firms. As there is no direct information on capital costs available, in this study capital costs are estimated from fire-insurance values. We have also obtained sector-specific input and output price indexes. These price indexes were deflated using the consumer price index to obtain real-price indexes. The resulting (deflated) indexes were set to 1 in the first observation year 1976, i.e. $p_0 = 0$. This normalization implies a rescaling of X_{tj}^i (according to the relation (2)). That is, X_{tj}^i is output/factor quantity multiplied by an unknown constant. However, our model is invariant to this transformation: the only effect is an induced shift in the intercept of the first equation in (4).

Because use of fossil fuels show a great variance across sectors, we concentrate on fossil fuel-intensive sectors: Fish oil and fish meat (ISIC 31151), paper and paper products (ISIC 341), and mineral products (ISIC 369).

4.1 Estimation results

The firms were observed in the period 1976-95. Firms that entered the industry during the observation period were also included in the sample. Because of the VAR(1) structure, firms were sampled conditional on being operative in at least two periods. The interest rate was set to 7 per cent, while a planning horizon of $H = 20$ was chosen. Recall that

$$x_t^i = (\text{log-output, log-capital, log-labor, log-energy, log-materials})'$$

and

$$y_t^i = (\text{revenue, capital cost, labor cost, energy cost, material cost})'.$$

A principal components analysis of the y_t^i data was performed. The number of principal components used in the option value approximation (19) was decided according to the following rule: Start with the first principal component Z_{t1}, and add additional components until the corresponding β-parameter estimate in (22) is insignificant at the 95 per cent level. In all sectors, this procedure led to selection of two principal components.

Preliminary estimates of θ and ω were obtained by a two-stage estimation procedure: First the parameters θ of the VAR model were estimated by maximizing the corresponding *partial* pseudo log-likelihood, i.e. the likelihood obtained by considering the stopping times T_i as deterministic. Then

ω was estimated based on the logistic function (22), given the θ estimate from the first stage. This partial estimation procedure was followed by joint estimation based on the full pseudo likelihood. Computer programs were written in Gauss to carry out these tasks. The fit of our model is compared with a model with a myopic (Marshallian) exit rule based on current profit π_t^i (instead of Π_t^{*i}) and with no option value. Below our main results are reported for each of the sectors in turn.

Fish oil and fish meal

The estimates from this sector are based on 47 firms. We first look at the estimated time-series matrix-parameter $\widehat{\Phi}$:

$$\widehat{\Phi} = \begin{bmatrix} .36 & .10 & .18 & .11 & .25 \\ .08 & .93 & -.11 & .06 & -.01 \\ .08 & .04 & .83 & .06 & -.05 \\ -.30 & .12 & .36 & .74 & .19 \\ .20 & .11 & .29 & -.05 & .57 \end{bmatrix}$$

$\widehat{\Phi}$ has two complex eigenvalues, indicating cyclical variations in the adjustment toward equilibrium. Furthermore, $\widehat{\Phi}$ has one eigenvalue which is not significantly different from 1: the estimate is 0.99, with a standard deviation of 0.01.[13] This suggests that there may be four, rather than five, cointegration vectors in the system $(x_t^{i\,\prime}, p_t{}')'$ (see the discussion at the end of Section 2).

The estimates of the logistic parameters are shown in Table 1. Recall that β_0 is the constant term; β_1 and β_2 are the coefficients of the first and second principal components of y_t^i; γ is the coefficient of the random effect v^i; and α is the coefficient of the net present value Π_t^{i*}. We note that $\widehat{\alpha}$ is negative and highly significant: high net present value leads to a low exit probability.

The first two principal components are primarily linear combinations of labor and capital. This finding is replicated in the other two sectors. It is plausible to interpret these principal components as proxies for firm size. The rationale may be that, for a given level of expected profitability, a firm's ability to take advantage of unexpected positive shocks is increasing in the stock of physical and human capital. As required, the estimated option values are positive for all firms in all observation years. The random effect v^i, however, was not found to have any significant impact on the variation in option values across firms: $\widehat{\gamma} \approx 0$.

In this, as well as in the other two sectors, we find that the cross-sectional variance in random effects is relatively small (although $\widehat{\Upsilon}$ is significantly different from 0). For example, $\mathrm{tr}(\widehat{\Upsilon}\widehat{\Upsilon}')/\mathrm{tr}(\widehat{\Sigma}) \approx 0.01$. This does not, however, imply that the variation in v^i is unimportant in order to explain heterogeneity among firms: In all sectors, we find evidence of either a unit root or that ideosyncratic shocks have a

[13] In order to make statistical inferences which are consistent with the pseudo likelihood approach, standard deviations were calculated using the "sandwich matrix" formula; see e.g. Gourieroux and Montfort (1995), p. 237.

Parameter	Estimate	St.dev.
β_0	-3.80	0.36
β_1	-0.35	0.05
β_2	-0.83	0.35
γ	-0.00	0.12
α	-4.30	1.31

Table 1: Fish oil and fish meal: Estimated logistic parameters

high degree of persistence. In the presence of a unit root, v^i will give rise to a linear trend in x_t^i, and thus its impact on firm heterogeneity is magnified over time.

The size of the sector has been dramatically reduced over time: only 7 of the 43 firms operating in 1977 were active in 1996, and there were only 4 entries between 1976 and 1995. Figure 1a (i.e. the upper chart of Figure 1) shows observed exit frequencies versus average estimated exit probabilities among *all* firms for each year in the period 1977-95. Figure 1a also depicts, for each year, the average exit probability of (i) firms that survived throughout the *whole* observation period ("survivors"), and (ii) firms that exited during the observation period ("exiting firms"). The model appears to discriminate well between survivors and exiting firms: the average exit probability of survivors is substantially lower than that of exiting firms.

In order to assess the fit of the model, we have calculated McFadden's ρ^2, which in this sector equals 0.21. For comparison, partial estimation led to $\rho^2 = 0.20$, while ρ^2 is only 0.01 in the myopic model. The ability of the model to discriminate between firms is also indicated in Figure 1b (i.e. the lower chart of Figure 1). This figure depicts the distribution of exit probabilities of exiting and non-exiting "firm-years" over the whole observation period: a "firm-year" is a firm observed in one year, and it is classified as "non-exiting" if the firm is operating throughout the next year and "exiting" if it exits (at the beginning) of the next year.[14] As anticipated, the distribution of exiting firm-years has the heaviest tail. In particular, the average exit probabilities of exiting and non-exiting firm-years are 17 per cent and 8 per cent, respectively.

Paper and paper products

The sample consists of 119 firms. The estimated time-series matrix-parameter $\widehat{\Phi}$ is

$$\widehat{\Phi} = \begin{bmatrix} .83 & .01 & .18 & -.04 & .03 \\ .12 & .73 & .15 & .03 & -.04 \\ .16 & .00 & .81 & -.01 & .01 \\ .14 & .04 & .11 & .90 & -.18 \\ .29 & .00 & .24 & -.07 & .57 \end{bmatrix}$$

[14] To take one example, suppose a firm is active in 1976 and 1977, but exits (early) in 1978. Its estimated exit probability for 1976 is assigned to the distribution of non-exiting firm-years, whereas the estimated exit probability for 1977 is assigned to the distribution of exiting firm-years.

Parameter	Estimate	St.dev.
β_0	-7.84	1.18
β_1	-2.22	0.46
β_2	-4.73	1.72
γ	-0.08	0.16
α	-0.94	0.29

Table 2: Paper and paper products: Estimated logistic parameters

The eigenvalues of $\widehat{\Phi}$ lie in the interval 0.52 to 0.99, with a standard deviation of 0.01 for the highest eigenvalue. Thus, again we find evidence for four cointegrating vectors. The estimates of the logistic parameters are shown in Table 2. In particular, all β-parameters are clearly significant. Also α is significant, although the estimate is much lower than in the previous sector. We found that $\rho^2 = 0.23$ (partial estimation led to $\rho^2 = 0.22$). This is dramatically higher than for the myopic model, where $\rho^2 = 0.01$.

In this sector, there were 107 firms in 1977, and 46 of them remained active in 1996. In addition, there were 12 entries during the observation period. Only 3 of these firms were still active in 1996. Estimated versus observed annual exit probabilities are depicted in Figure 2a. The figure shows a reduction in the average exit probability of all firms from 1977 to 1995. This seems to be mainly due to self-selection among the 1977 firms. That is, the most profitable firms have survived.

Figure 2b shows the distribution of exit probabilities for exiting versus non-exiting firm-years. Again we find a marked difference in the shape of these distributions: the average exit probabilities in the two distributions are 15 per cent and 5 per cent, respectively.

Mineral products

This sample consists of 246 firms. We obtained the time-series estimates

$$\widehat{\Phi} = \begin{bmatrix} .41 & .03 & .17 & .01 & .27 \\ .03 & .64 & .17 & .08 & .06 \\ .00 & .02 & .91 & .02 & .02 \\ -.06 & .08 & .13 & .78 & .04 \\ .10 & .02 & .04 & -.01 & .80 \end{bmatrix}$$

The eigenvalues of $\widehat{\Phi}$ are in the range of 0.35 to 0.96, where the latter estimate has a standard deviation of 0.02. The estimates of the logistic parameters are shown in Table 3. All β-parameters, γ, and α are clearly significant. The average exit probabilities of exiting and non-exiting firm-years are 9 per cent and 4 per cent, respectively (see Figure 3b).

In this sector, 91 of the 176 active firms in 1977 survived until 1996. Moreover, there were 70 entries, of which 29 firms were still operating in 1996. On the other hand, Figure 3a shows peaks in

Parameter	Estimate	St.dev.
β_0	-3.29	0.14
β_1	-1.07	0.20
β_2	-2.27	0.49
γ	-0.60	0.14
α	-5.96	1.22

Table 3: Mineral products: Estimated logistic parameters

the exit rates around 1983-84 and, in particular, during 1987-92. These periods are characterized by low output prices, primarily due to a low demand for mineral products from the construction sector.

Figure 4 shows, for each year, industry-wide adjustment of output and labor input to changes in prices, i.e. the first and third component of Γp_t (see equation (4)). The vertical axis measures (approximately) percentage changes relative to 1977 in aggregate output and aggregate labor use. As seen from Figure 4, both components are low in the recessions (when observed exit rates are high). However, the model does not capture the dramatic increase in exits during 1987-92: there is only a moderate increase in aggregate estimated exit probabilities compared with previous years. This may be the main reason for a low ρ^2; $\rho^2 = 0.11$. Moreover, with partial estimates $\rho^2 = 0.07$, whereas in the myopic model $\rho^2 = 0.04$.

4.2 Policy simulations

Increased concern for the greenhouse effect has led to a number of local and global initiatives to cut emissions of carbon dioxide (CO_2) and other harmful greenhouse gases. In Norway, for example, a state-appointed committee recently released a report on greenhouse gas quotas for the Norwegian economy (NOU 2000:1). The committee's recommendations were partly based on results obtained from an application of the present model. In this section, we explain how the model can be applied to examine the impact on firm exit when firms are forced to buy quotas. For each firm, the amount of acquired quotas corresponds to the emission of greenhouse gases following from combustion of fossil fuels (emissions are expressed in units of CO_2).

Economic theory suggests that firms respond to a higher price of fossil fuels (due to the quotas) through (*i*) substitution between energy goods away from fossil fuels, (*ii*) substitution away from energy, and (*iii*) adjustment of the scale of production. In order to build these effects into our model, detailed information about e.g. substitution elasticities and scale elasticities is required. Unfortunately, there are few studies on Norwegian manufacturing sectors that provide adequate information on any of these parameters. Given the limited information, we focus on only one type of response, namely substitution between different energy goods. In order to study this effect, we need to impose more structure on the underlying production process.

Firstly, as in Section 2, the production function must be weakly separable in the major input categories capital, labor, energy, and materials. Secondly, we assume that the energy aggregate is homothetic in its inputs fossil fuels and electricity. These assumptions imply that the firm's choice of factor inputs can be decomposed into two steps: First choose the (relative) composition of the energy aggregate, and then choose the optimal level of capital, labor, energy, and materials; see Denny and Fuss (1977).

Let F_t^i denote the total quantity of fossil fuels, and let f_t^i denote the quantity of fossil fuels per unit of energy in firm i:

$$f_t^i = \frac{F_t^i}{X_{tE}^i}.$$

If we mark costs that include acquisition of quotas with *, the cost of consuming X_{tE}^i under a quota system is

$$y_{tE}^{i*} = (\upsilon_t c f_t^i + p_{tE}) X_{tE}^i = \tau_t^i y_{tE}^i,$$

where υ_t is the quota price at time t, and c is emission of CO_2 per unit of fossil fuels, and

$$\tau_t^i = \frac{\upsilon_t c f_t^i + p_{tE}}{p_{tE}} \left(= \frac{\upsilon_t CO_2^i + y_{tE}^i}{y_{tE}^i} \right) \tag{23}$$

is the price of energy with quotas relative to the price of energy without quotas. In equation (23), we have used the notation CO_2^i to denote CO_2 emissions from firm i at time t. Note that $\ln \tau_t^i$ can be interpreted as the relative change in the unit cost of energy due to quotas.

We now investigate the relation between τ_t^i and f_t^i. For simplicity, we drop i and t and use the notation $\tau(f)$ to denote τ as a function of f, and $f(\cdot)$ to denote f as a function of the price of fossil fuels. Let f^0 and f^1 be the level of f before and after the quotas, respectively, and let p_F^i be the price of fossil fuels. We then have the following first-order approximation:

$$\begin{aligned} \ln \tau &\approx \ln \tau(f^0) + \frac{\lambda c}{\lambda c f + p_E} f\prime(p_F) \lambda c \\ &= \ln \tau(f^0) + \mathrm{El}_{p_F} F \frac{\lambda CO_2}{\lambda CO_2 + y_E} \frac{\lambda CO_2}{y_F} \end{aligned} \tag{24}$$

The first term, $\ln \tau(f^0)$, denotes the relative change in the unit cost of energy due to the quotas if there is no substitution within the energy aggregate. The second term, which is non-positive, modifies this effect by taking into account the effect of re-optimizing the mix of the energy aggregate. The magnitude of the latter effect is the product of (i) the conditional elasticity of demand for fossil fuels with respect to its own price (total amount of energy is fixed), (ii) the quota payment relative to the total costs of energy, and (iii) the ratio between the quota payment and the cost of fossil fuels before the quotas. Note that since the optimal f_t^i minimizes $\upsilon_t c f_t^i + p_{tE}$ (the average costs of energy with quotas), the price of energy p_{tE} will also increase. But this is a second-order effect under the assumption of optimization, and is assumed to be negligible.

We now turn to how to model expectations about future τ_t^i. Because we have no data on υ_t, we simply assume that

$$E_t\{\tau_{t+k}^i\} = \tau_t^i, \tag{25}$$

that is, τ_t^i is a *martingale* with respect to the information set of the firm at t. Hence, the firm expects the percentage change in the tax payment per unit of energy and the price of energy, P_{tE}, to be equal.

In the presence of a quota price $\upsilon_t > 0$, we define the net present value as

$$\Pi_t^{i*}(e) = \iota'(\sum_{k=1}^{e} \delta^k E_t\{y_{t+k}^{*i}\})$$

where

$$y_{tj}^{*i} = \begin{cases} \tau_t^i y_{tE}^i & j = E \\ y_{tj}^i & j \neq E \end{cases}$$

In this study, we set the price of quotas equal to USD 25 per tonne of CO_2 in 1996. This number corresponds roughly to the quota price in a competitive market containing all signatories of the Kyoto agreement (averaged over the estimates of nine studies); see Weyant and Hill (1999). Moreover, the constrained direct demand elasticity for fossil fuels (-0.34) is taken from Pindyck (1979). When firms are forced to buy carbon quotas, profits fall and hence the firm-specific exit probabilities increase. Table 4 shows the results for our three sectors for the 1996 exit probabilities. Because firms in these sectors are small and primarily sell their products in the world market, all prices can be taken to be exogenous. We study the impact of quotas under two different assumptions: (i) no substitution between energy inputs, and (ii) substitution between energy inputs.

As seen from Table 4, the impact on 1996 exit probabilities of imposing carbon quotas is small. For fish oil and fish meal, the average exit probability increases from 4.3 per cent to 4.7 per cent, and equals 4.6 per cent when firms substitute between energy inputs.[15] A closer examination reveals that the highest increase in firm-specific exit probabilities is around 1 percentage point. For paper and paper products, as well as mineral products, the rise in average exit probability (following from the quotas) is as low as 0.1 percentage points. Also in these two sectors, most firms experience a minor increase in the exit probability, and no increase exceeds 1.5 percentage points.

[15] Assume that all annual firm-specific exit probabilities are constant over time (and equal to the 1996 values), and that quotas are not imposed. After 10 years, we would expect that 72 per cent of the active firms in 1996 are still operating. With quotas (and no substitution), the corresponding number is 70 per cent.

Sector	No quotas	Quotas – no subst.	Quotas – subst.
Fish oil and fish meal	4.3	4.7	4.6
Paper and paper products	2.5	2.6	2.5
Mineral products	4.4	4.5	4.4

Table 4: Average annual exit probabilities in 1996 (in per cent) when the price of quotas is USD 25 per tonne of CO_2 in 1996

5 Concluding remarks

The purpose of the present paper has been to apply stochastic dynamic programming in order to develop an econometric model for firm exit. According to stochastic dynamic programming, the value of continuing production can be written as a sum of two terms: (i) the net present value of being committed to stay operative until a fixed future date (where the time to exit is chosen according to the Stock and Wise approach), and (ii) the value of keeping the exit option alive. Due to obstacles in applying stochastic dynamic programming, we approximate the option value by a simple function of its determinants. By combining the exit decision rule derived from stochastic dynamic programming and a stochastic model for firm profitability, and moreover approximating the value of the exit option, we derive probability distributions of firm exit. The econometric model is estimated by pseudo likelihood methods, using panel data from the Norwegian Manufacturing Statistics. In general, the data are explained well by the model. The applicability of the model is then illustrated by assessing to what extent quotas on CO_2 emissions increase firm exits in three carbon-intensive manufacturing sectors. We find that a quota price of USD 25 per tonne CO_2 (in 1996) raises most firm-specific exit probabilities only slightly.

A Estimation

We will now turn to estimation of the logistic parameters $\omega = (\beta_0, \beta, \gamma, \alpha)$, the time-series parameters $\theta = (\Phi, \Gamma, \Sigma, \mu, \Upsilon)$ in (4)-(5), and the parameters ψ in (6). For notational simplicity, assume that all firms enter the sample at $t = 0$ (the general case is a straightforward extension).

All probability statements will henceforth be conditional on prices and x_0^i, although this conditioning is mostly suppressed in the notation.

A.1 The pseudo likelihood function

Conditional on x_0^i, the observed data on firm i consist of $\{(z_t^i, x_t^i); t = 1, ..., T_i\}$. T_i is the exit time or, if there is no observed exit, $T_i = T$ – the last possible observation year. This means that T_i is a realization of a random variable, say, τ_i. We will now establish the pseudo likelihood (conditional on the x_0^i's) as a function of (ω, θ, ψ).

Define the stopped stochastic process $X_{\tau_i}^i = \{x_1^i, ..., x_{\tau_i}^i\}$, and let $X_{T_i}^i = \{x_1^i, ..., x_{T_i}^i\}$ for a *fixed* index T_i. Furthermore, let $f(X_{T_i}^i)$ denote the probability density of $X_{T_i}^i$. Since, for Borel sets B_t, $[x_1^i \in B_1, ..., x_{\tau_i}^i \in B_{\tau_i}, \tau_i = T_i]$ and $[x_1^i \in B_1, ..., x_{T_i}^i \in B_{T_i}, \tau_i = T_i]$ describe the same event, the probability density of $X_{\tau_i}^i$ can be factorized as $P(\tau_i = T_i | X_{T_i}^i) f(X_{T_i}^i)$.

Let $\widehat{P}_\omega(\cdot;\theta)$ be the approximate exit probability distribution defined in (22). Furthermore, let $f(v^i | X_{T_i}^i; (\theta, \psi))$ be the density of v^i conditional on $X_{T_i}^i$, and $f(X_{T_i}^i; (\theta, \psi))$ the marginal density of $X_{T_i}^i$. Then the joint pseudo log-likelihood function $\widehat{l}(\omega, \theta)$ becomes

$$\widehat{l}(\omega, \theta, \psi) = \sum_{i=1}^{N} \widehat{l}^i(\omega, \theta, \psi), \tag{26}$$

where N is the number of firms, and

$$\begin{aligned} \widehat{l}^i(\omega, \theta, \psi) = \ln \int \prod_{t=1}^{T_i} \widehat{P}_\omega(z_t^i | v^i, x_t^i, p_t, z_{t-1}^i = 0; \theta) f(v^i | X_{T_i}^i; (\theta, \psi)) dv^i \\ + \ln f(X_{T_i}^i; (\theta, \psi)). \end{aligned} \tag{27}$$

The first term on the right-hand side of (27) corresponds to $\ln \widehat{P}_\omega(\tau_i = T_i | X_{T_i}^i)$, while the second term is the log of the marginal density of $X_{T_i}^i$. Note that if $\tau_i = T$, then $z_{T_i} = 0$ and $z_{T_i} = 1$ are both possible: the observation period may end with exit or non-exit. On the other hand, $\tau_i < T$ implies $z_{T_i} = 1$.

We are now in a position where we can prove the claim in Section 2, that is, the initial distribution of v^i (conditional on x_0^i) affects only the maximum (pseudo) likelihood estimates of the *other* primitives of the model through $f(v^i | X_{T_i}^i; (\theta, \psi))$. Let $\rho_\psi(v^i)$ denote some class of positive inital distributions of v^i indexed by ψ. That is, ψ may have a different meaning than in the familiy (6). Furthermore, let $(\widehat{\theta}(\psi), \widehat{\omega}(\psi))$ denote maximum pseudo likelihood estimates for fixed ψ. Then we have the following result:

Proposition 1 *$(\widehat{\theta}(\psi), \widehat{\omega}(\psi))$ depends on the initial distribution $\rho_\psi(v^i)$ only through $f(v^i | X_{T_i}^i; (\theta, \psi))$.*

Proof. Since $\ln f(X_{T_i}^i; (\theta, \psi)) = \ln f(X_{T_i}^i | v^i; \theta) + \ln \rho_\psi(v^i) - \ln f(v^i | X_{T_i}^i; (\theta, \psi)$, we have, for *any* fixed θ_0 and ψ_0,

$$\begin{aligned} \ln f(X_{T_i}^i; (\theta, \psi)) \quad = \quad & \int \ln f(X_{T_i}^i | v^i; \theta) f(v^i | X_{T_i}^i; (\theta_0, \psi_0)) dv^i + \int \ln \rho_\psi(v^i) f(v^i | X_{T_i}^i; (\theta_0, \psi_0)) dv^i \\ & - \int \ln f(v^i | X_{T_i}^i; (\theta, \psi) f(v^i | X_{T_i}^i; (\theta_0, \psi_0)) dv^i \end{aligned} \tag{28}$$

(note that this is the kind of representation which is used in the EM algorithm[16]). It is seen that only

[16] See e.g. Tanner (1993), p. 43.

the first and third terms on the right-hand side of (28) depend on θ, and only the latter term depends on ψ, and then only through $f(v^i|X^i_{T_i};(\theta,\psi)$. ■

A.2 The estimation algorithm

A natural estimation strategy would be to maximize the pseudo likelihood with respect to the unknown parameters. However, this is a rather complex problem. First, it is not clear how we should evaluate the distribution of $v^i|X^i_{T_i}$ which appears in (27). Furthermore, $\widehat{P}_\omega(z^i_t|v^i,x^i_t,p_t,z^i_{t-1}=0;\theta)$ depends in a complex way on the parameters θ through the net present value $\Pi^i_t(v^i;\theta)$.

To be able to solve the estimation problem, we shall start by obtaining a useful representation of the density $f(X^i_{T_i};(\theta,\psi))$. Let

$$R^i_t=\Upsilon v^i+\varepsilon^i_t. \tag{29}$$

The first equation in (4) can then be written

$$x^i_t=\mu+\Phi x^i_{t-1}+\Gamma p_t+R^i_t. \tag{30}$$

As originally proposed in Raknerud (1999), define *orthogonal* residuals

$$\widetilde{R}^i_t=\begin{cases}\sqrt{\frac{t}{t+1}}(R^i_{t+1}-\frac{1}{t}\sum_{v=1}^t R^i_v) & t=1,..,T_i-1\\ \frac{1}{T_i}\sum_{v=1}^{T_i}R^i_v & t=T_i\end{cases} \tag{31}$$

It is easily seen from (5)-(6) and (29)-(31) that

$$f(X^i_{T_i};(\theta,\psi))\propto\prod_{t=1}^{T_i-1}f(\widetilde{R}^i_t;\theta)\times f(\widetilde{R}^i_{T_i};(\theta,\psi))$$

where

$$\widetilde{R}^i_t=\begin{cases}\sqrt{\frac{t}{t+1}}\Big((x^i_{t+1}-\frac{1}{t}\sum_{v=1}^t x^i_v)-\Phi(x^i_t-\frac{1}{t}\sum_{v=1}^t x^i_{v-1})-\Gamma(p_{t+1}-\frac{1}{t}\sum_{v=1}^t p_v)\Big) & t=1,..,T_i-1\\ \frac{1}{T_i}\sum_{v=1}^{T_i}(x^i_v-\mu-\Phi x^i_{v-1}-\Gamma p_v) & t=T_i,\end{cases}$$

and

$$\widetilde{R}^i_t\sim\begin{cases}\mathcal{N}(0,\Sigma) & t=1,...,T_i-1\\ \mathcal{N}(\Upsilon\psi'(x^i_0-E\{x^i_0\}),\Upsilon\Upsilon'+T_i^{-1}\Sigma) & t=T_i.\end{cases}$$

Hence

$$\begin{aligned}\widehat{l^i}(\omega,\theta)&=\ln\int\prod_{t=1}^{T_i}\widehat{P}_\omega(z^i_t|v^i,x^i_t,p_t,z^i_{t-1}=0;\theta)f(v^i|X^i_{T_i};(\theta,\psi))dv^i\\&\quad+\sum_{t=1}^{T_i-1}\ln f(\widetilde{R}^i_t;\theta)+\ln f(\widetilde{R}^i_{T_i};(\theta,\psi)).\end{aligned} \tag{32}$$

Since only $\widetilde{R}^i_T$ depends on v^i,

$$f(v^i|X^i_{T_i};(\theta,\psi))=f(v^i|\widetilde{R}^i_{T_i};(\theta,\psi)). \tag{33}$$

That is, $\widetilde{R}^i_{T_i}$ is prediction sufficient for v^i (see Bjørnstad (1996)); there is no information about v^i in $\widetilde{R}^i_t$ for $t < T_i$. From (29) and (31),

$$\widetilde{R}^i_{T_i}|v^i \sim \mathcal{N}(\Upsilon v^i, T_i^{-1}\Sigma) \tag{34}$$

and we can establish the following result:

Proposition 2 *The distribution of v^i given $X^i_{T_i}$ is given by*

$$v^i|X^i_{T_i} \sim \mathcal{N}\left((\Upsilon'\Sigma^{-1}\Upsilon + T_i^{-1})^{-1}\left(\Upsilon'\Sigma^{-1}\widetilde{R}^i_{T_i} + T_i^{-1}\psi'(x^i_0 - E\{x^i_0\})\right), (T_i\Upsilon'\Sigma^{-1}\Upsilon + 1)^{-1}\right). \tag{35}$$

Proof. The proposition is an immediate consequence of (6), (33), (34), and Bayes's theorem. ■

We note that for given (θ, ψ), maximization of (32) with respect to the logistic parameters ω is quite easy. The integral can be evaluated numerically or by Monte Carlo integration. Moreover, because the variance in v^i conditional on the time-series data $X^i_{T_i}$ is of the order $O(T_i^{-1})$, as seen from (35), a very accurate approximation to the integral can be obtained by replacing v^i with its conditional mean in (35).

To estimate ω, θ, and ψ, we propose a two-step procedure. In the first step, simple preliminary estimates $(\widetilde{\omega}, \widetilde{\theta}, \widetilde{\psi})$ are obtained. These are then used as starting values when maximizing (32) jointly with respect to (ω, θ, ψ).

It is well known that for any stopping time $\widetilde{\tau}_i$ which is adapted to I^o_t,

$$E\left\{\begin{array}{c} \frac{\partial \ln f(x^i_1,..,x^i_{\widetilde{\tau}_i};(\theta,\psi))}{\partial\theta} \\ \frac{\partial \ln f(x^i_1,..,x^i_{\widetilde{\tau}_i};(\theta,\psi))}{\partial\psi} \end{array}\right\} = 0, \tag{36}$$

that is, (θ, ψ) maximizes $E\{\ln f(X^i_{T_i}; (\theta, \psi))\}$ (see Hall and Heyde (1989), p. 155). Unfortunately (36) does not hold for τ_i, because τ_i depends on v^i. However, if little information about v^i is revealed by observing $\{z^i_1, ..., z^i_{T_i}\}$, (36) might hold approximately for the stopping time τ_i. This motivates the initial estimator $(\widetilde{\theta}, \widetilde{\psi})$ defined by

$$\frac{1}{N}\sum_{i=1}^{N}\left[\begin{array}{c} \frac{\partial \ln f(x^i_1,..,x^i_{T_i};(\widetilde{\theta},\widetilde{\psi}))}{\partial\theta} \\ \frac{\partial \ln f(x^i_1,..,x^i_{T_i};(\widetilde{\theta},\widetilde{\psi}))}{\partial\psi} \end{array}\right] = 0,$$

which is found by maximizing $\sum_{i=1}^{N}\left(\sum_{t=1}^{T_i-1}\ln f(\widetilde{R}^i_t;\theta) + \ln f(\widetilde{R}^i_{T_i};(\theta,\psi))\right)$. A preliminary estimator of ω, say, $\widetilde{\omega}$, can then be obtained by logistic regression:

$$\widetilde{\omega} = \arg\max_{\omega}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\ln \widehat{P}_\omega\left(z^i_t|E\{v^i|X^i_{T_i};(\widetilde{\theta},\widetilde{\psi})\}, x^i_t, p_t, z^i_{t-1} = 0\,;\widetilde{\theta}\right). \tag{37}$$

That is, v^i in (32) is replaced by its conditional expectation (evaluated at $\theta = \widetilde{\theta}$), and the resulting approximation to $\widehat{l}^i(\omega, \widetilde{\theta})$ is maximized with respect to ω.

In the second step, we obtain *full* maximum pseudo likelihood estimators using $(\widetilde{\omega}, \widetilde{\theta}, \widetilde{\psi})$ as staring values. Although the "logistic part" of (32), $\ln \int \prod_{t=1}^{T_i} \widehat{P}_\omega(z_t^i | v^i, x_t^i, p_t, z_{t-1}^i = 0\,; \theta) f(v^i | X_{T_i}^i; (\theta, \psi)) dv^i$, is a complicated function of θ, it is smooth. For example, derivatives of $\Pi_t^{i*}(v^i; \theta)$ with respect to θ can be obtained by applying the chain rule on (10)-(11) with $e = e_t^*$ and then recursively obtaining derivatives of $E_t(x_{t+k}^i + p_{t+k})$ and $Var_t(x_{t+k}^i + p_{t+k})$ from (4).[17] Although $\Pi_t^{i*}(v^i; \theta)$ is a non-smooth function of θ – because the discrete choice e_t^* depends on θ – averaging over v^i "smooths out" the discontinuities in $\frac{\partial \Pi_t^{i*}(v^i;\theta)}{\partial \theta}$.

Our data set confirms that the final estimates, $(\widehat{\omega}, \widehat{\theta}, \widehat{\psi})$, are close to the initial estimates, $(\widetilde{\omega}, \widetilde{\theta}, \widetilde{\psi})$, and that the method converges quite quickly. Visual inspection of the log-likelihood in orthogonal directions (corresponding to the eigenvectors of the estimated covariance matrix) confirms that we find global maximizers.

References

Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. John Wiley, New York.

Bjørnstad, J.F. (1996). "On the Generalization of the Likelihood Function and the Likelihood Principle". *Journal of the American Statistical Association*, 91, 791-806.

Das, S. (1992). "A Micro Econometric Model of Capital Utilization and Retirement: The Case of the Cement Industry". *Review of Economic Studies*, 59, 277-297.

Denny, M. and M. Fuss (1977). "The Use of Approximation Analysis to Test for Separability and the Existence of Consistent Aggregates". *American Economic Review*, 67, 404-418.

Dixit, A.K. and R.S. Pindyck (1994). *Investment under Uncertainty*. Princeton University Press, Princeton.

Engle, R.F. and C.W.J. Granger (1987). "Co-integration and Error Correction: Representation, Estimation and Testing". *Econometrica*, 55, 251-276.

Gourieroux, C. and A. Montfort (1995). *Statistics and Econometric Models*. Cambridge University Press, Cambridge.

Hall, P. and C.C. Heyde (1989). *Martingale limit theory and its applications*. Academic Press, Boston.

Klette, T.J. and Z. Griliches (2000). "Empirical Patterns of Firm Growth and R&D-Investment: A Quality Ladder Model Interpretation". *Economic Journal*, 110, 363-387.

[17] See Lütkepohl (1996), Ch. 10, for a survey of the relevant matrix differentiation rules.

Lehman, E.L. (1983). *Theory of Point Estimation.* John Wiley, New York.

Lütkepohl, H. (1996). *Handbook of Matrices.* John Wiley, New York.

Marshall, A. (1966). *Principles of Economics.* 8th edition. London. Macmillan.

NOU 2000:1. *Quotas for Greenhouse Gases.* Governement of Norway, Oslo.

Olley, G.S. and A. Pakes (1996). "The Dynamics of Productivity in the Telecommunications Equipment Industry". *Econometrica*, 64, 1263-1297.

Pakes, A. and R. Ericson (1998). "Empirical Implications of Alternative Models of Firm Dynamics". *Journal of Economic Theory,* 79, 1-45.

Pesaran, M.H. and R. Smith (1995). "The Role of Theory in Econometrics". *Journal of Econometrics*, 67, 61-79.

Pindyck, R.S. (1979). "Interfuel Substitution and the Industrial Demand for Energy: An International Comparison". *The Review of Economics and Statistics*, LXI, 169-179.

Raknerud, A. (1999). "Maximum Likelihood Estimation of VAR Models for Panel Data with Dynamic Latent Components" in *Dynamic Models for Paneld Data. Methods and Applications.* Ph.D. dissertation, University of Oslo, Dept. of Economics.

Rust, J. (1987). "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurchner", *Econometrica*, 55, 999-1033.

Rust, J. (1994). "Structural Estimation of Markov Decision Processes" in R.F. Engle and D.L. McFadden (eds.): *Handbook of Econometrics.* Vol. 4. North-Holland, Amsterdam.

Rust, J. (1997). "Using Randomization to Break the Curse of Dimensionality". *Econometrica*, 65, 487-516.

Stern S. (1997). "Approximate solutions to stochasic programs". *Econometric Theory*, 13, 392-405.

Stock, J.H. and D.A. Wise (1990). "Pensions, the Option Value of Work, and Retirement". *Econometrica*, 58, 1351-1380.

Summers, L.H. (1987). "Investment Incentives and the Discounting of Depreciation Allowances" in M. Feldstein (ed.): *The Effects of Taxation on Capital Accumulation*, Chicago University Press, Chicago.

Sutton, J. (1997). “Gibrat’s legacy”. *Journal of Economic Literature*, 35, 40-59.

Tanner, M.A. (1993). *Tools for Statistical Inference*. Springer, New York.

Weyant, J.P. and J. Hill (1999). “Introduction and Overview” in J. P. Weyand (ed.): *The Costs of the Kyoto Protocol: A Multi-Model Evaluation. The Energy Journal*, special issue.

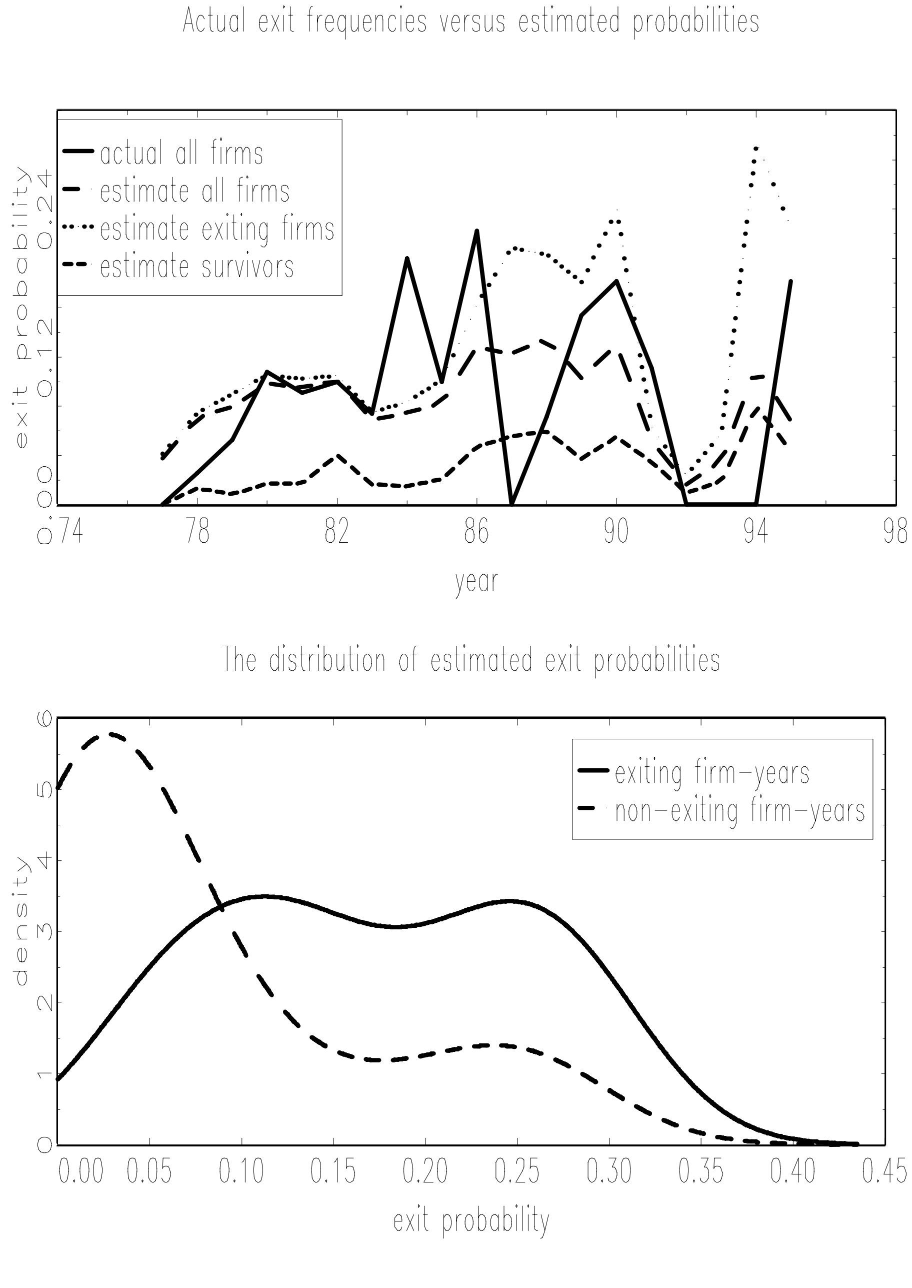


Figure 1: Fish oil and fish meal: Comparison of exit probabilities between exiting and non-exiting firms.

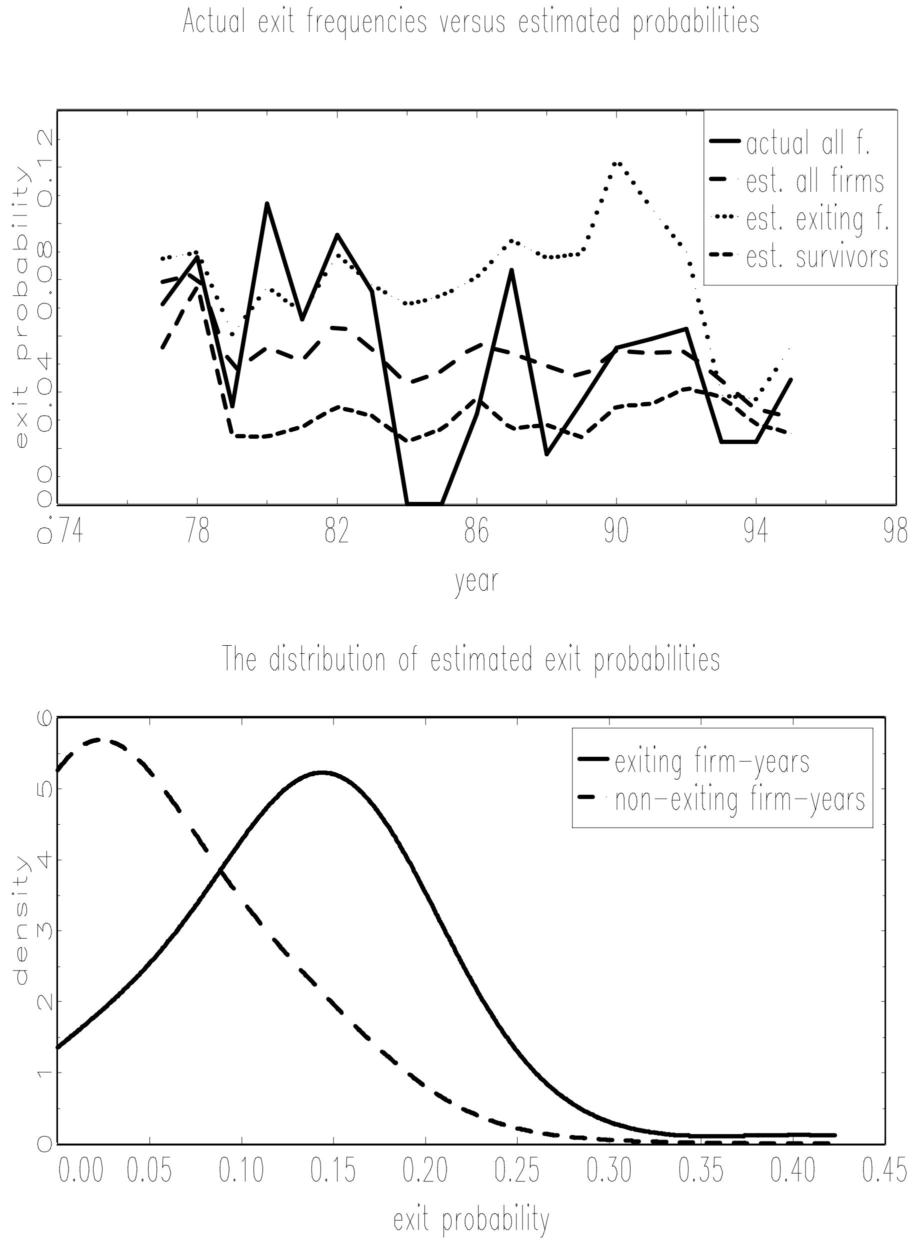
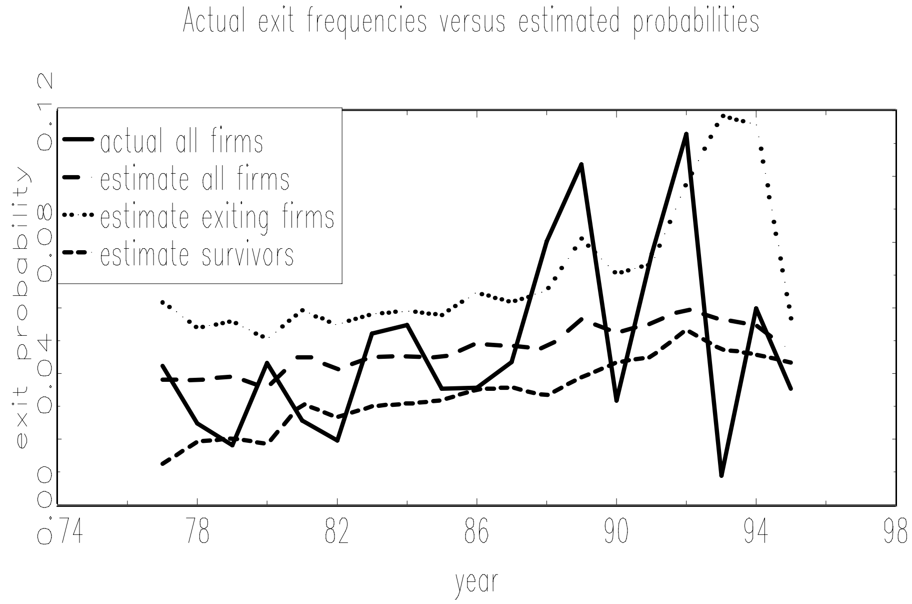


Figure 2: Paper and paper products: Comparison of exit probabilities between exiting and non-exiting firms.



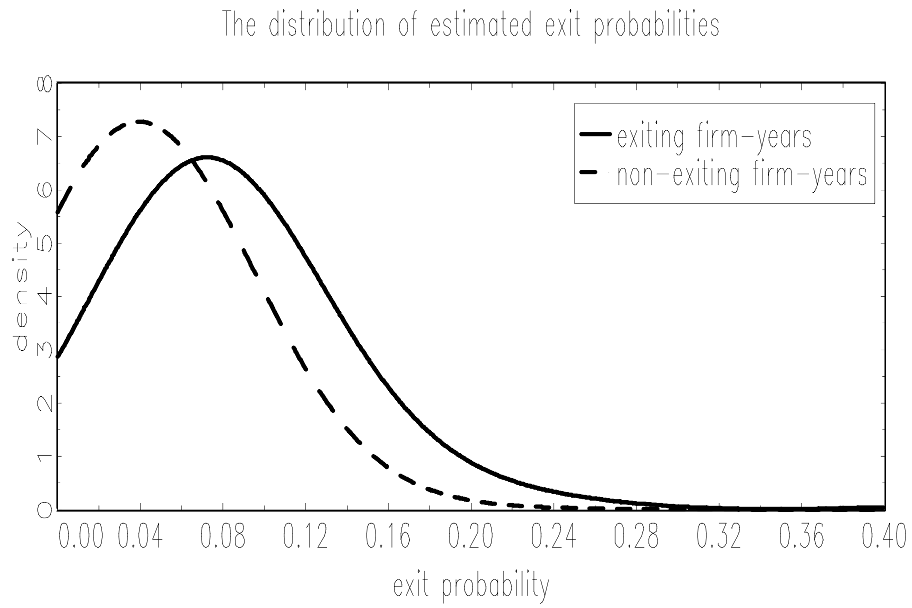


Figure 3: Mineral products: Comparison of exit probabilities between exiting and non-exiting firms.

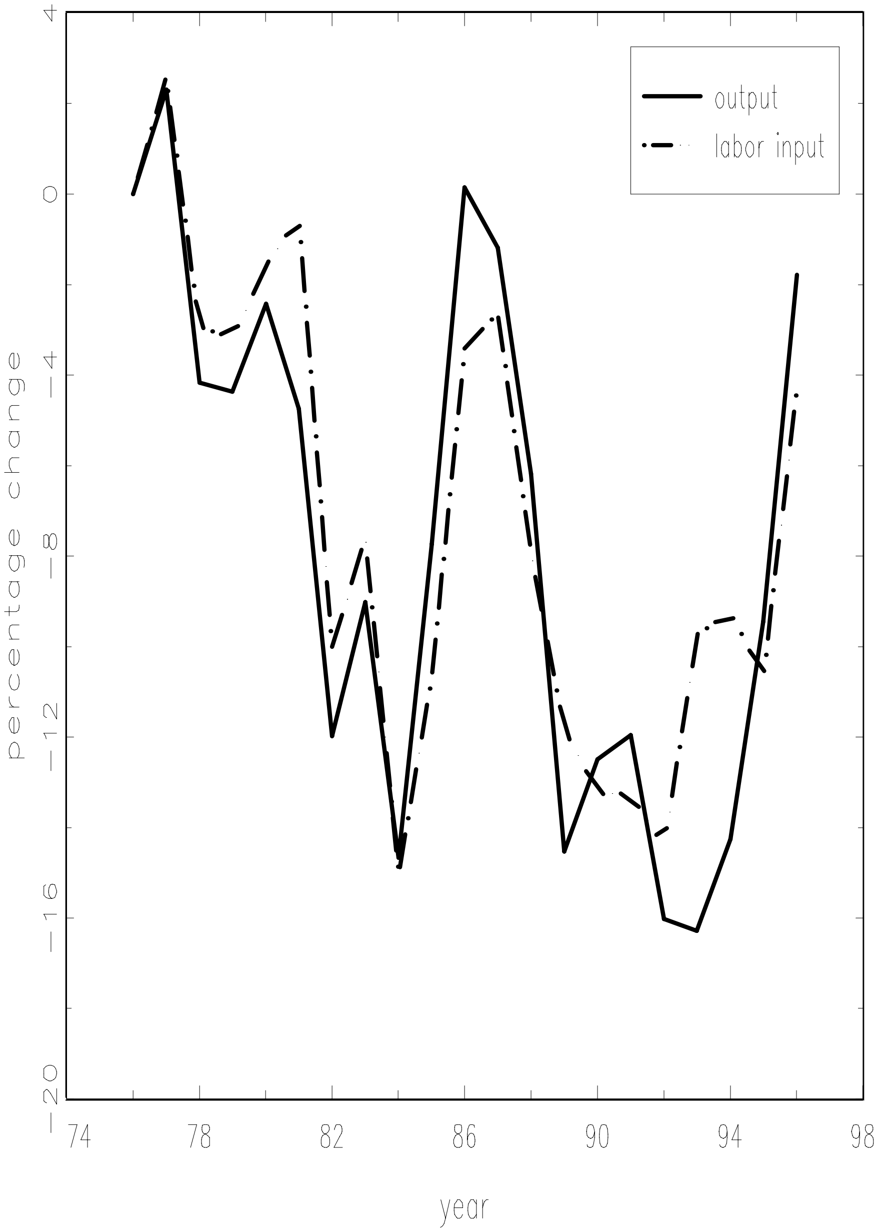


Figure 4: Mineral products: Fluctuations in aggregate activity level due to changes in prices.