Journal of Monetary Economics 56 (2009) 184–199

Contents lists available at ScienceDirect

Journal of Monetary Economics

journal homepage: www.elsevier.com/locate/jme

The scarring effect of recessions ☆

Min Ouyang

Department of Economics, 3151 Social Science Plaza B, Irvine, CA 92697-5100, USA

ARTICLE INFO

Article history:
Received 22 January 2007
Received in revised form
19 December 2008
Accepted 19 December 2008
Available online 20 January 2009

JEL classification:
E32
L16
C61

Keywords:
Cleansing effect
Scarring effect
Creative destruction
Learning
Demand shocks

ABSTRACT

According to the conventional view, recessions improve resource allocation by driving out less productive firms. This paper posits an additional scarring effect: recessions impede the developments of potentially superior firms by destroying them during their infancy. A model is developed to capture both the cleansing and the scarring effects. A key ingredient of the model is that idiosyncratic productivity is not directly observable, but can be learned over time. When calibrated with statistics on entry, exit and productivity differentials, the model suggests that the scarring effect dominates the cleansing effect, and gives rise to lower average productivity during recessions.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

How do recessions affect resource allocation? Economists have long studied this question. Schumpeter (1934) advanced the concept of cleansing: recessions eliminate outdated techniques and obsolescent products, and thus free resources for more productive uses. This idea has been revived during the last decade in an assortment of theoretical work such as Caballero and Hammour (1994, 1996), Hall (2000), and Mortensen and Pissarides (1994). In recent years, however, researchers have begun to explore alternative ways in which recessions might influence allocation. Barlevy (2002, 2003) posits adverse effects caused by credit-market frictions or on-the-job search that offset some or all of the cleansing effect. Related empirical investigations suggest that recessions affect allocation through numerous channels, some with characteristics consistent with cleansing and others consistent with negative effects of the alternatives.[1]

Three existing empirical findings have pointed to an unexplored channel through which recessions can affect allocation. First, although businesses' deaths do surge during recessions, the failing ones are not always the least productive. For

☆ The author is indebted to Robert King, an anonymous referee, Richard Rogerson, and John Shea for their detailed comments. Ricardo Caballero, Gadi Barlevy, Mark Gertler, John Haltiwanger, Hugo Hopenhayn, Micheal Pries, Vincenzo Quadrini, John Rust, Dmitriy Stolyarov, and Daniel Vincent provided generous advice. The author also thank John Haltiwanger for providing the gross job flows data. Seminar participants at UMD, Yale, USC, UC-Irvine, Kansas City Fed, Federal Reserve of San Francisco, Federal Reserve of Cleveland, McGill, Queen's, UC-San Diego, UC-Davis, provided helpful discussions. Technical appendix is provided at the ScienceDirect and the author's homepage. All errors are the author's.

E-mail address: mouyang@uci.edu
URL: http://www.socsci.uci.edu/~mouyang

[1] See Davis et al. (1999) and Barlevy (2002, 2003).

0304-3932/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jmoneco.2008.12.014

example, Baden-Fuller (1989) examines British recessions during the 1980s and finds that many closing firms were more profitable than the surviving ones. Second, most of the businesses fail at a very young age. According to Dunne et al. (1989), over 75 percent of the exiting plants in the U.S. manufacturing sector aged five years or less (Table 1, p. 676). Third and most importantly, recessions *disproportionately* affect businesses in their early years of operation. This is shown in Fig. 1, which plots the quarterly exit rate of the U.S. manufacturing plants across three age categories. Apparently, infant plants suffer the most during recessions. For example, in the second quarter of 1984, the exit rate jumped from 1.35% to 3.42% for plants aged one year or less, and from 0.78% to 0.87% for plants aged between one and nine years; but, for plants aged 10 years or more, it rose from 0.35% to 0.37% only.

The disproportionate deaths among infant businesses, this paper argues, should play an important role in determining the allocative effect of economic downturns. Infant businesses tend to appear unproductive in the short run, but have the potential to reveal high productivity in the future. Recessions that destroy infant businesses scar the economy, by preventing new and innovative businesses from reaching their full potential. This scarring effect offsets the conventional cleansing effect, although both effects take place through the exit of unprofitable firms. Accordingly, the overall impact of recessions on resource allocation depends upon the relative magnitudes of these two competing effects—cleansing and scarring.

To understand the scarring effect, consider the life cycle of a firm. A firm usually starts without fully knowing its own quality. Uncertainty may come from the unobserved talent of the manager, unknown appeal for the product, or unpredictable profitability of a retail location. As the firm operates, realized revenue signals its true quality: high revenue indicates that it is productive and encourages continuing in operation; low revenue implies otherwise. The longer a firm operates, the more it learns about its true quality. Therefore, potentially good firms—those that do not yet know they are good—must be relatively young. During recessions, profitability declines in general so that a firm cannot bear to learn as *long* as during good times. A potentially good firm that would have survived during good times, might thus exit during recessions *before* it learns. At the industry level, the exit of potentially good firms reduces the proportion of good firms at present times, as well as in the future because fewer potentially good firms are left to learn. The reduced proportion of good firms lowers average productivity, which is defined in this paper as a scarring effect.

The above story reflects the spirit of learning, theoretically proposed by Jovonavoic (1982) and empirically promoted by Caves (1998) and Foster et al. (2008) as a powerful tool to understand firm turnover. In this paper, a simplified learning mechanism from Pries (2004) is combined with the vintage framework of Caballero and Hammour (1994) to capture cleansing and scarring theoretically. The model decomposes firm productivity into two components—vintage and unobservable idiosyncratic productivity—so that an industry's average productivity is determined by the distribution of firms across both dimensions. The idiosyncratic productivity is not directly observable, but can be learned over time. Demand variations serve as the source of economic fluctuations. Lower demand reduces profitability in general, so that firms exit younger. Younger exit ages direct, on the one hand, resources to younger and more productive vintages, causing a

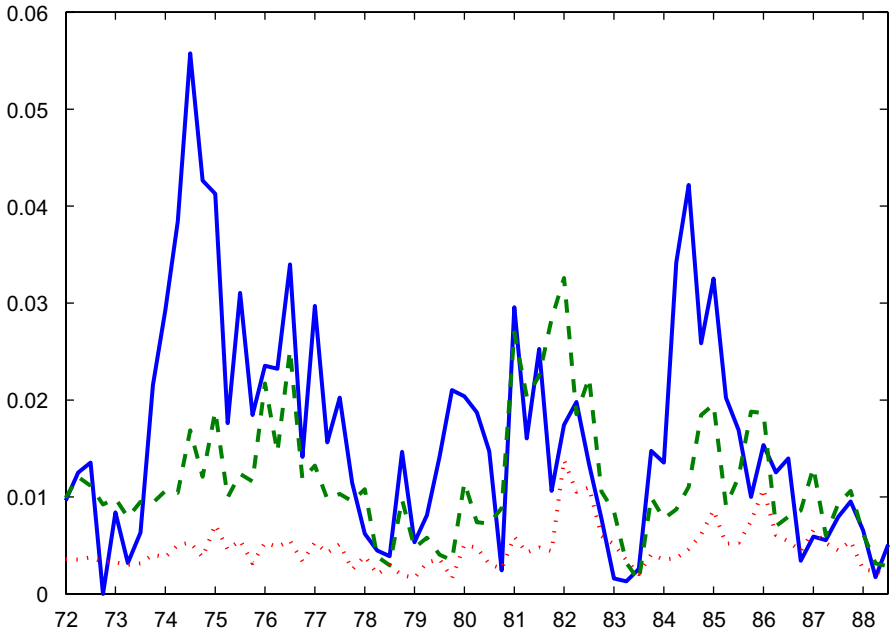

Fig. 1. Quarterly plant exit rates in the U.S. manufacturing sector (1972–1988). Dotted line represents the exit rate for plants aged ten years or older; dashed line represents the exit rate of plants aged between one and ten years; solid line represents that for plants younger than a year. All exit rates are employment-weighted. Data source: gross job flows data compiled by Davis et al. (1996).

cleansing effect that raises average productivity; on the other hand, they *truncate* the learning process that lead resources toward firms with higher idiosyncratic productivity, creating a scarring effect that pulls down average productivity. Hence, recessions cause two competing effects—cleansing and scarring. The question then becomes, which effect dominates?

The paper turns to data to evaluate the scarring effect quantitatively. The model is calibrated to statistics on entry, exit, and cohort productivity differentials observed in the U.S. manufacturing sector. When applied to stochastic demand shocks, the calibrated model suggests that the scarring effect dominates the cleansing effect and generates lower average productivity during recessions.

The rest of the paper is organized as follows. Section 2 lays out the model. The cleansing and scarring effects are motivated in Section 3 with comparative static exercises. Section 4 applies the model to stochastic demand shocks, confirms that the cleansing and scarring effects carry over, and studies their quantitative implications. Section 5 concludes.

2. A renovating industry with learning

Consider an industry where labor and capital combine in fixed proportions to produce a homogenous output. Firms that enter in different periods coexist, each characterized by two components: vintage and idiosyncratic productivity.

A firm's vintage is given by an exogenous technological progress that drives the industry's leading technology to grow at a constant rate $\gamma>0$. With A_t as the leading technology in period t, $A_{t+1}=A_t(1+\gamma)$. Firms that enter in period t adopt A_t. Only entrants have access to new technology; incumbents cannot retool. With firm age defined as the number of periods that a firm has survived through, the vintage of a firm of age a in period t equals $A_t(1+\gamma)^{-a}$.

At entry, each firm is endowed with idiosyncratic productivity θ. This idiosyncratic productivity can be the talent of the manager as in Lucas (1978) or, alternatively, the location of the store, the organizational structure of the production process or its fitness to the embodied technology. The key assumption regarding θ is that its value, although fixed at the time of entry, is not directly observable.

A firm hires one worker. The period-t output of a firm of age a and with idiosyncratic productivity θ is

$$q_t(a,\theta)=A_t(1+\gamma)^{-a}x_t, \tag{1}$$

where $x_t=\theta+\varepsilon_t$. x_t captures the influence of θ on output masked by an independent transitory shock ε_t. With the wage rate normalized as one and the output price denoted as P_t, this firm's period-t profit is

$$\pi_t(a,\theta)=P_tA_t(1+\gamma)^{-a}(\theta+\varepsilon_t)-1. \tag{2}$$

Both $q_t(a,\theta)$ and $\pi_t(a,\theta)$ are directly observable. A firm knows its vintage and can infer the value of x_t by observing output or revenue. Given knowledge of the distribution of ε_t, a firm uses the value of x_t to learn about θ.

2.1. "All-or-nothing" learning

Firms attempt to resolve the uncertainty about θ to decide whether to continue or terminate production. Following Pries (2004), we model an "all-or-nothing" learning process, assuming only two values of θ: θ_g for a good firm and θ_b for a bad firm. Moreover, ε_t is distributed uniformly on $[-\omega,\omega]$, so that a good firm will have x_t each period as a random draw from a uniform distribution over $[\theta_g-\omega,\theta_g+\omega]$, and a bad firm will have it drawn over $[\theta_b-\omega,\theta_b+w]$. θ_g, θ_b and ω satisfy $0<\theta_b-\omega<\theta_g-\omega<\theta_b+\omega<\theta_g+\omega$.

Therefore, an observation of x_t within $(\theta_b+\omega,\theta_g+\omega]$ indicates a firm has good idiosyncratic productivity; conversely, an observation of x_t within $[\theta_b-\omega,\theta_g-\omega)$ tells that it has bad idiosyncratic productivity. However, an x_t within $[\theta_g-\omega,\theta_b+\omega]$ reveals nothing, because the probabilities of falling in this range as a good firm and as a bad firm both equal to $(2\omega+\theta_b-\theta_g)/2\omega$.

This all-or-nothing learning process simplifies the model considerably, as it gives only three values for θ^e (the expected θ). Correspondingly, there are three groups of firms in the industry: good firms with $\theta^e=\theta_g$, bad firms with $\theta^e=\theta_b$, and "unsure firms" with $\theta^e=\theta_u$, the prior mean of θ. The probability of true idiosyncratic productivity being revealed every period is $p\equiv(\theta_g-\theta_b)/2\omega$. The unconditional probability of $\theta=\theta_g$ is exogenous and equals φ. A firm enters the market as unsure; thereafter, every period it stays unsure with probability $1-p$, learns it is good with probability $p\varphi$, and learns it is bad with probability $p(1-\varphi)$. Therefore, the evolution of θ^e from the time of entry is a Markov process with values $(\theta_g,\theta_u,\theta_b)$, an initial probability distribution $(0,1,0)$, a transition matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ p\varphi & 1-p & p(1-\varphi) \\ 0 & 0 & 1 \end{pmatrix}, \tag{3}$$

and a limiting probability distribution as a goes to ∞, $(\varphi,0,1-\varphi)$. If firms were to live forever, eventually all uncertainty would be resolved, as enough information would be provided to reveal each firm's true idiosyncratic productivity.

Suppose that each entering cohort consists of a continuum of firms, so that the law of large numbers applies. Then φ and p are not only probabilities, but also the fractions of firms with $\theta=\theta_g$ in a entering cohort and of firms each period that learn their true idiosyncratic productivity. *Ignoring firm exit for now*, the fractions of good firms, of unsure firms, and of bad

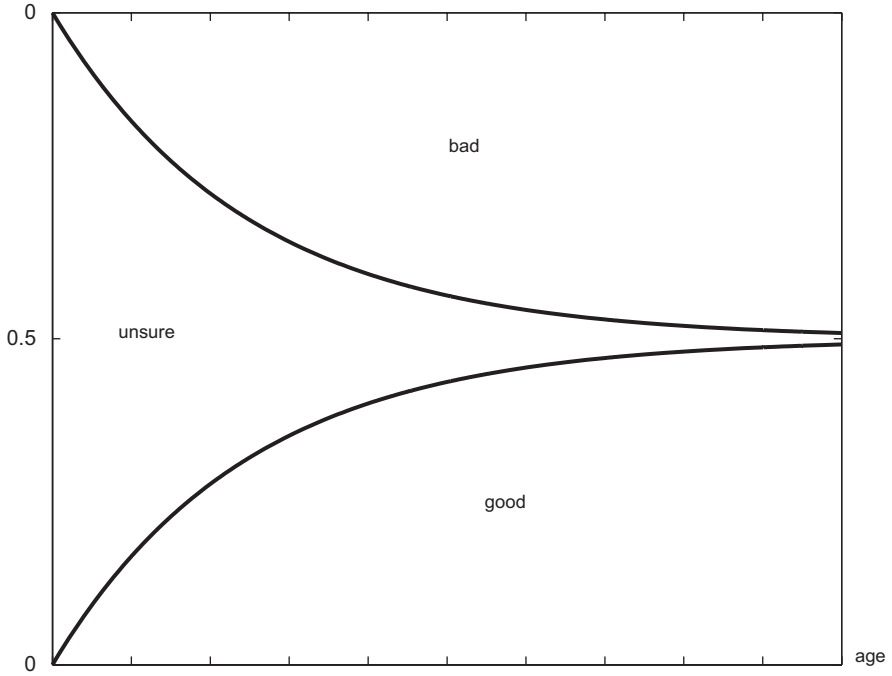

Fig. 2. Dynamics of a birth cohort with all-or-nothing learning: the distance between the bottom curve and the bottom axis measures the density of firms with $\theta^e = \theta_g$; the distance between the top curve and the top axis measures the density of firms with $\theta^e = \theta_b$; the distance between the two curves measures the density of firms with $\theta^e = \theta_u$. This figure is generated assuming a prior fraction of good firms of 0.5 and a learning pace of 0.02.

firms in a cohort of age a are

$$(\varphi - \varphi(1-p)^a,\ (1-p)^a, (1-\varphi) - (1-\varphi)(1-p)^a). \tag{4}$$

Fig. 2 plots the evolution of firm distribution within a birth cohort under all-or-nothing learning. The horizontal axis depicts the age of a cohort *across time*. Apparently, the fractions of firms that know their true idiosyncratic productivity, whether good or bad, grow as a cohort ages. Moreover, the two learning curves that denote the dynamics of factions of good firms and bad firms are both *concave*. This is the decreasing property of marginal learning captured by Jovonavoic (1982): the marginal learning effect decreases as a firm ages, which, in this model, is reflected as the decline in the marginal number of learners as a cohort ages. The convenient feature of all-or-nothing learning is that, on the one hand, the firm-level learning occurs suddenly, which allows for a easy track of the cross-section firm distribution, while, on the other hand, cohort-level learning takes place gradually as in a more standard learning process.

However, there is more that Fig. 2 can tell. Let the horizontal axis to depict the cross-sectional distribution of firm ages at any instant, then Fig. 2 captures the firm distribution across ages and idiosyncratic productivity in an industry that features constant entry but no exit. In this industry, cohorts continuously enter in the same size and experience the same dynamics as they age, so that, at any one time, different life stages of different birth cohorts overlap, giving rise to the firm distribution in Fig. 2. Under this interpretation, Fig. 2 indicates that older cohorts contain fewer unsure firms, as they have lived longer and learned more.

2.2. The recursive competitive equilibrium

Within each period, the sequence of events occurs as follows. First, entry and exit take place after firms observe the aggregate state of the industry. Second, each surviving firm pays a fixed operating cost to produce. Third, the output price is realized. Fourth, firms observe revenue and update their beliefs. Then, another period begins.

With this setup, this subsection considers a *recursive competitive equilibrium* definition, which includes as a key element the law of motion of the aggregate state of the industry. The aggregate state is (F, D). F denotes the firm distribution across vintages and idiosyncratic productivity. In F, the element that measures the number of firms of age a and with belief θ^e is $f(\theta^e, a)$. D is an observable exogenous demand parameter. The law of motion for D is exogenous, described by D's transition matrix. The law of motion for F is endogenous and denoted as $H : F' = H(F, D)$. The sequence of events implies that H captures the influence of entry, exit and learning on firm distribution.

Three assumptions characterize this industry equilibrium: firm rationality, free entry, and competitive pricing.

2.2.1. Firm rationality

Firms are forward-looking price takers and profit maximizers. They predict current and future profitability to make decisions on entry or exit. The relevant state variables for a firm are its vintage, its belief about its true idiosyncratic productivity, and the aggregate state (F, D). Let $V(\theta^e, a; F, D)$ to be the expected value, for a firm of age a and with belief θ^e, of staying in operation for one more period and optimizing afterward when aggregate state is (F, D). Then V satisfies

$$V(\theta^e, a; F, D) = E[\pi(\theta^e, a)|F, D] + \beta E[\max(0, V(\theta^{e\prime}, a+1; F', D'))|F, D], \tag{5}$$

subject to $F' = H(F, D)$, the law of motion for D, and the law of motion for θ^e driven by all-or-nothing learning. Since firms enter as unsure, the expected value of entry is $V(\theta_u, 0; F, D)$. According to the firm rationality condition, entry occurs only when $V(\theta_u, 0; F, D) > 0$; a firm of age a and with belief θ^e exits if and only if $V(\theta^e, a; F, D) < 0$.[2]

2.2.2. Free entry

Under the free entry condition, new firms can enter at any instant as long as they bear an entry cost c. This entry cost can be the cost of establishing a particular location, purchasing capital stock, or finding a qualified manager. Let $f(\theta_u, 0; F, D)$ to be the size of an entering cohort with aggregate state (F, D). The entry cost is assumed to be a linear function of the entry size:

$$c = c_0 + c_1 f(\theta_u, 0; F, D), \quad c_0 > 0 \text{ and } c_1 > 0. \tag{6}$$

$c_1 > 0$ suggests that the entry cost increases in the entry size. This can arise from a limited amount of land available for production sites or, alternatively, an upward-sloping supply curve for the industry's capital stock. Goolsbee (1998) provides supporting evidence for this assumption, showing that higher investment demand raises the equipment prices. Goolsbee's finding suggests that, as more firms enter, capital price rises with capital demand, so that entry becomes more costly.

Therefore, new firms keep entering as long as the expected value of entry exceeds the cost of entry. At the same time, the entry cost keeps rising until reaching $V(\theta_u, 0; F, D)$. At this point, entry stops, and

$$V(\theta_u, 0; F, D) = c_0 + c_1 f(\theta_u, 0; F, D). \tag{7}$$

2.2.3. Competitive pricing

The output price is determined by

$$P(F, D) = \frac{D}{Q(F, D)}, \tag{8}$$

where Q is total industry output; it equals the sum of production over heterogeneous firms. Recall that, according to the sequence of events, production takes place *after* entry and exit. Let F' to be the updated firm distribution after entry and exit, and $f(\theta^e, a)'$ to be the element of F' that measures the number of firms of age a and with belief θ^e.[3] Applying (1) gives

$$Q(F, D) = Q(F') = \sum_a \sum_{\theta^e} A(1+\gamma)^{-a} \theta^e f(\theta^e, a)', \tag{9}$$

where D is an exogenous demand parameter that captures the influence of demand fluctuations on an industry's production profitability. In reality, such demand variations can arise from changes in consumers' taste on an industry's production goods or, alternatively, productivity shocks of a down-stream industry that demands this industry's output as one of its inputs. In this model, D equals industry total revenue, and it is the exogenous fluctuations in D that introduce cycles to the industry. Higher D implies higher P, which encourages entry while reduces exit, so that Q rises. Conversely, lower D causes less entry but more exit, and Q falls.

With the conditions of firm rationality, free entry, and competitive pricing, the following definition is established:

Definition. A *recursive competitive equilibrium* is a law of motion H, a value function V, and a pricing function P such that (i) V solves the firm's optimization; (ii) P satisfies (8); and (iii) H is generated by the decision rules suggested by V and the appropriate summing-up of entry, exit and learning.

An additional assumption is made to simplify the model:

Assumption. Given values for other parameters, the value of θ_b is so low that $V(\theta_b, a; F, D)$ remains negative for any a and any (F, D).

[2] Caballero and Hammour (1994) assume myopic exit behavior—a firm exits as long as its current profit drops below zero—by modeling a deterministic and smooth demand sequence, under which the exit behavior of a forward-looking firm is similar to that of a myopic firm. In contrast, demand follows a two-state Markov process in our model so that firm exit decisions must be forward-looking. The forward-looking exit behavior incorporates the value of waiting. When demand is low, a firm may choose to stay even if its current profit has dropped below zero, as it realizes a probability of future demand recovery.

[3] Q is contributed by the *expected* output $A(1+\gamma)^{-a}\theta^e$ instead of the realized output $A(1+\gamma)^{-a}(\theta+\varepsilon)$. This is because, with a continuum of firms in each birth cohort, the law of large numbers applies, so that the production noises ε and the expectation errors cancel out and the sum of realized output equals the sum of expected output.

Under this assumption, bad firms always exit. Thus, at any one time, there are only two types of firms in operation—unsure and good.

3. Cleansing and scarring

Section 2 shows that the firm distribution F enters the model as a state variable, which makes it difficult to characterize the industrial dynamics in response to stochastic demand shocks. However, it is generally true that, if shocks are sufficiently persistent, the effects of temporary changes in response to transitory shocks are similar to those induced by permanent shocks. Therefore, this section uses comparative static exercises on the steady-state equilibrium to motivate the cleansing and scarring effects. The next section will turn to a numerical analysis of the model to confirm that the two effects carry over with stochastic demand shocks.

3.1. Steady state

A steady state is a recursive competitive equilibrium with time-invariant aggregate states. In a steady state, D is and is *perceived* as time-invariant, $D' = D$; F is also time-invariant, $F = H(F, D)$. Because H is generated by entry, exit, and learning, a steady state must feature time-invariant entry and exit for $F = H(F, D)$ to hold. Thus, a steady state can be summarized by $\{f(0), \bar{a}_g, \bar{a}_u\}$: $f(0)$ is the time-invariant entry size; $\bar{a}_g$ is the maximum age for good firms; and $\bar{a}_u$ is the maximum age for unsure firms. Proposition 1 establishes the existence of a unique steady-state equilibrium for any D.

Proposition 1. *With constant D, there exists a unique time-invariant $\{f(0), \bar{a}_g, \bar{a}_u\}$ that satisfies the conditions of firm rationality, free entry and competitive pricing.*

Detailed proof is provided in Appendix A. A key step of the proof is to combine the exit condition for unsure firms and that for good firms to get

$$\left(\frac{\theta_u}{\theta_g} + \frac{p\varphi\beta}{1+\gamma-\beta}\right)(1+\gamma)^{\bar{a}_g - \bar{a}_u} = 1 + \frac{p\varphi\beta}{1-\beta} - \frac{p\varphi\beta\gamma}{(1-\beta)(1+\gamma-\beta)}\beta^{\bar{a}_g - \bar{a}_u}. \tag{10}$$

The proof for Proposition 1 shows that, with $\theta_g > \theta_u$, (10) determines a unique value for $\bar{a}_g - \bar{a}_u$. Note that D does not enter (10), so that demand has no impact on $\bar{a}_g - \bar{a}_u$. With $\bar{a}_g - \bar{a}_u$ determined by (10) independently, $\bar{a}_u$ in the free entry condition and the competitive pricing condition can be replaced by $\bar{a}_g - (\bar{a}_g - \bar{a}_u)$. As a result, those two conditions jointly determine the values for $f(0)$ and $\bar{a}_g$.

Fig. 3 illustrates the steady-state firm distribution. Like Fig. 2, it can be interpreted in two ways. First, let the horizontal axis to depict the cohort age *across time*, then Fig. 3 displays the steady-state life-cycle dynamics of a representative cohort. A cohort enters as unsure in a measure of $f(0)$. As it ages, bad firms exit and the cohort size declines; good firms stay and the density of good firms grows. At age $\bar{a}_u$, all unsure firms exit with their vintage too old to survive as unsure, but good

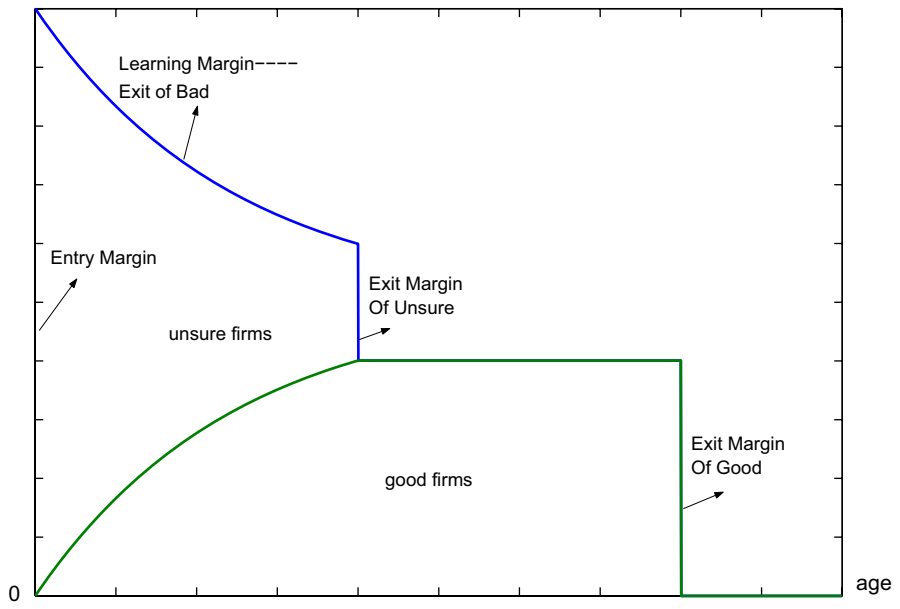

Fig. 3. The steady-state firm distribution and the entry and exit margins.

firms can stay. After $\overline{a_u}$, learning stops, the cohort contains good firms only, and its size remains constant. Good firms live until $\overline{a_g}$. The vintage after $\overline{a_g}$ is too old even for good firms to survive.

Second, let the horizontal axis in Fig. 3 to depict the cohort age *cross section*. Then Fig. 3 displays the steady-state firm distribution across ages and idiosyncratic productivity at any one time. Firms of different ages coexist. Because older cohorts have lived longer and learned more, their sizes are smaller and their densities of good firms are higher. Cohorts older than $\overline{a_u}$ are of the same size and contain good firms only. No cohort is older than $\overline{a_g}$.

Also note that, despite its time-invariant structure, the industry experiences continuous entry and exit at the steady state. From a pure accounting point of view, there are three margins for resources to flow in this industry: the entry margin, the exit margins of good firms and unsure firms, and the learning margin. At the entry margin, new vintages enter; at the exit margins, old vintages leave. This introduces a force of creative destruction that replaces old vintages with new vintages. At the learning margin, bad firms leave. This gives rise to a learning force that keeps good firms and drives out bad firms. Because of creative destruction, average productivity grows at the technological pace γ. Because of learning, the proportion of good firms is higher among older cohorts. The two forces—learning and creative destruction—together drive entry, exit, and the related productivity dynamics.

3.2. Comparative statics: cleansing and scarring

This subsection establishes that, across steady states, the model delivers the conventional cleansing effect, and an additional scarring effect. The two effects are formalized in Propositions 2 and 3.

Proposition 2. *In a steady-state equilibrium, the exit age for firms with a given θ^e is weakly increasing in demand.*

Detailed proof is provided in Appendix A. Proposition 2 suggests that firms with any belief live longer in a high-demand steady state. Put intuitively, lower demand reduces price, so that some firms that are viable when demand is high become not viable when demand is low.

If this story carries over when D fluctuates stochastically, then the model delivers the conventional cleansing effect, in which average firm age falls so that the average vintage becomes younger and more productive. However, once learning is allowed, the firm distribution across the other dimension—idiosyncratic productivity—must be considered. With only two values for true idiosyncratic productivity, good and bad, this distribution can be summarized as the fraction good firms. The next proposition establishes how demand affects this ratio.

Proposition 3. *In a steady state equilibrium, the fraction of good firms is weakly increasing in demand.*

Detailed proof is presented in Appendix A. The steady-state industry fraction of good firms including both known and yet unknown, denoted as l_g, can be shown as

$$l_g = 1 - \frac{(1-\varphi)}{\dfrac{p\varphi(\overline{a_u}+1)}{1-(1-p)^{\overline{a_u}+1}} + (1-\varphi) + p\varphi(\overline{a_g}-\overline{a_u})}. \tag{11}$$

There are only two endogenous variables in (11): $\overline{a_u}$ and $(\overline{a_g}-\overline{a_u})$. Since $(\overline{a_g}-\overline{a_u})$ is independent of D according to (10), $d(l_g)/d(D) = (d(l_g)/d(\overline{a_u}))\, d(\overline{a_u})/d(D)$. Appendix A shows that $d(l_g)/d(\overline{a_u}) \geqslant 0$, which, together with $d(\overline{a_u})/d(D) \geqslant 0$ established by Proposition 2, implies $d(l_g)/d(D) \geqslant 0$. Put intuitively, demand affects the fraction of good firms ($d(l_g)/d(D)$) through its impact on the exit age for unsure firms ($d(\overline{a_u})/d(D)$). To understand this result further, consider Fig. 4.

Fig. 4 displays the firm distribution across vintages and idiosyncratic productivity at a high-demand steady state and that at a low-demand steady state. Because the entry size scales the sizes of all age cohorts at a steady state, in Fig. 4 the entry sizes of both steady states are normalized as one. Fig. 4 shows that, corresponding to a lower demand, the two exit margins shift to the left, creating a cleansing effect that clears out oldest vintages. However, the leftward shift of the unsure exit margin also reduces the number of *older good firms*. The latter effect, shown as the shaded area in Fig. 4, is the scarring effect of recessions.

The scarring effect stems from learning. New entrants begin unsure of their idiosyncratic productivity, although a proportion φ are truly good. Firms learn their true idiosyncratic productivity over time. If firms could live forever, then all the potentially good firms would eventually realize their true idiosyncratic productivity. However, a finite life span of unsure firms implies that, if potentially good firms do not learn before $\overline{a}_u$, they exit at $\overline{a}_u$ and thus forever lose the chance to learn. Therefore, $\overline{a}_u$ represents not only unsure firms' exit age, but also the number of learning opportunities available in a firm's life time. A lower $\overline{a_u}$ gives potentially good firms less time to learn, so that the number of good firms in operation after age $\overline{a_u}$ is reduced.

Hence, the industry suffers from uncertainty: firms that exit at age $\overline{a_u}$ include some that remain unsure but are potentially good. The number of potentially good firms that exit at $\overline{a_u}$ depends on the size of the exit margin for unsure firms, which, in turn, is determined by $\overline{a_u}$. Lower demand truncates learning by reducing $\overline{a}_u$. Consequently, more potentially good firms exit at $\overline{a}_u$, fewer good firms become old, and the proportion of good firms for the entire industry declines.

To summarize from Propositions 2 and 3, a low-demand steady state features a better average vintage, yet a lower proportion of good firms. If these results carry over with stochastic demand shocks, then recessions will have both a

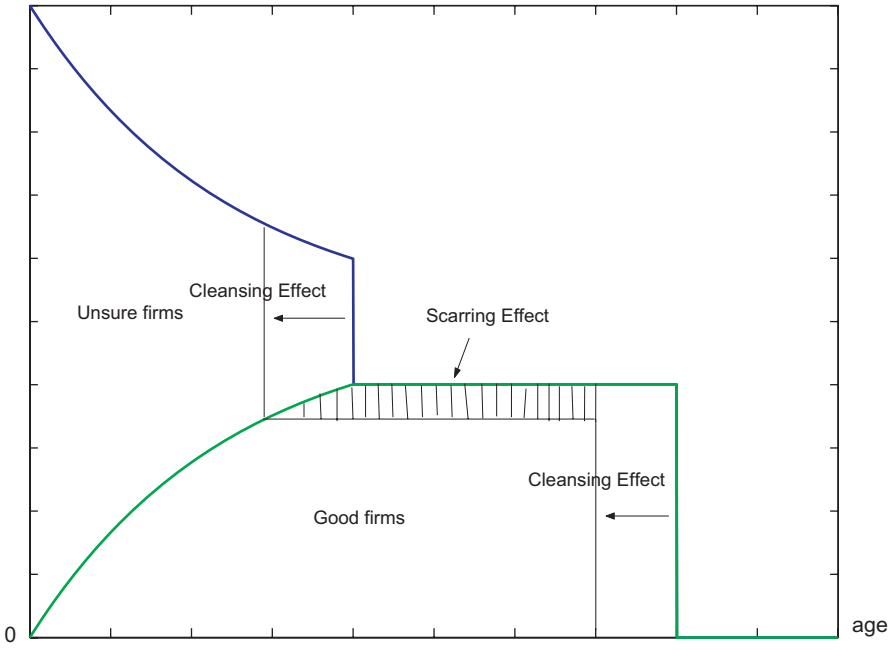

Fig. 4. Cleansing and scarring.

conventional cleansing effect that raises average vintage, and a scarring effect that lowers average idiosyncratic productivity. As suggested by $d(l_g)/d(D) = (d(l_g)/d(\overline{a_u}))\, d(\overline{a_u})/d(D)$, the two effects are directly related to each other: it is the cleansing effect ($d(\overline{a_u})/d(D)$) that truncates learning and prevents more potentially good firms from realizing their true idiosyncratic productivity, which causes the industry fraction of good firms to decline ($d(l_g)/d(\overline{a_u})$).

When moving beyond steady states and allowing for stochastic demand shocks, the intuition behind "cleansing and scarring" still carries over. Again, consider Fig. 4. When demand drops, the exit margins shift to the left, so that the cleansing effect takes place immediately. The scarring effect, however, occurs both instantaneously and gradually. At the onset of a recession, the fraction of good firms drops immediately, due to the shift of the exit margin for good firms that clear out oldest cohorts that contain good firms only—this is an "instantaneous scarring" effect. As the recession persists, another "lasting scarring" effect will follow. Note that, at the onset of a recession, the group of firms already in the shaded area in Fig. 4 would choose to *stay*, knowing their true idiosyncratic productivity to be good. These old good firms leave gradually as the recession persists, as there vintages grow more and more unproductive compared to other firms in operation. This creates a lasting scarring effect: the reduced $\overline{a_u}$ allows fewer potentially good firms to survive past $\overline{a_u}$, so that the shaded area would eventually be left blank. In summary, the arrival of a recession "scars" the industry, and the "scar" deepens as the recession persists. Instantaneous and lasting scarring effects together capture the impact of recessions on the industry composition of idiosyncratic productivity.

3.3. Sensitivity analysis

Three modeling assumptions should be discussed to examine the robustness of the scarring effect.

3.3.1. Entry cost and entry size

One of the key assumptions in the model is that entry cost increases in entry size: $c_1 > 0$. Will the scarring effect carry over if entry cost is independent of entry size? If $c_1 = 0$, then the conditions of firm rationality, free entry, and competitive pricing that jointly determine $\{f(0), \overline{a}_g, \overline{a}_u\}$ will become fully recursive: $\overline{a}_g - \overline{a}_u$ is given by (10) independently; with $\overline{a}_u = \overline{a}_g - (\overline{a}_g - \overline{a}_u)$, the free entry condition determines $\overline{a}_g$; then the competitive pricing condition, where D enters, determines $f(0)$. Therefore, D impacts $f(0)$ only when $c_1 = 0$. A detailed proof is presented in Appendix A. This extreme case is described as "full insulation" in Caballero and Hammour (1994): when $c_1 = 0$, *fast* entry is costless, so that the entry size adjusts proportionally to changes in demand and the exit margins remain unchanged. Therefore, with $c_1 = 0$, demand variations are entirely reflected as entry fluctuations, the exit margins do not respond. Consequently, there would be neither cleansing nor scarring effects.

On the contrary, when $c_1 > 0$, $f(0)$ and $\overline{a}_g$ are jointly determined by the free entry condition and the competitive pricing condition, so that some of the demand variations are accommodated at the entry margin, while the rest are taken as the shifts of the exit margins. Therefore, entry and exit both fluctuate over the cycle. Apparently, data are consistent with the

case when $c_1 > 0$. For example, the business employment dynamics (BED) data show that the quarterly plant entry rate and exit rate display very similar volatility in the U.S. economy from 1992 to 2007: the ratio of the standard deviation of the entry rate over that of the exit rate equals 1.00 for the manufacturing sector, and 1.02 for the entire private sector.[4]

3.3.2. Productivity composition of entrants

In the model, the proportion of good firms among entrants is exogenous and independent of demand. But it is likely for demand to affect entrants' average productivity. This is modeled by Pries and Rogerson (2005), who allow for inspection of unobservable firm quality prior to entry in addition to learning after entry. In that case, recessions raise the entry threshold for expected idiosyncratic productivity, so that only more promising firms enter. This would drive up the average idiosyncratic productivity among entrants during recessions, which plays against the scarring effect.

However, it is hard to tell whether recessions would in fact lower or raise entrants' average productivity. During recessions, while firms do need to feel more optimistic about themselves to enter, it also becomes cheaper to rent land and easier to find a qualified manager. Thus, it is possible that recessions actually *lower* the productivity threshold for firms to enter, which would complement the scarring effect by reducing entrants' average productivity. This is shown by Davis et al. (1996), who find that jobs created during recessions tend to be short-lived, and by Bowlus (1995), who estimates that jobs created during recessions are usually from lower part of the wage distribution. Further evidence is provided by Jensen et al. (2001), who document that the average productivity among the U.S. manufacturing entrants dropped during the 1982 recessions (Table 1, p. 327).

3.3.3. More complicated learning

Under the all-or-nothing learning in the model, the noises are distributed uniformly, and the expected idiosyncratic productivity, θ^e, takes on only two values: θ_u and θ_g. This has greatly simplified the analysis so that the scarring effect can be motivated analytically. However, one must consider whether the scarring effect will carry over to more complicated learning. Suppose that the noises covering the true idiosyncratic productivity is distributed normally with mean zero and variance σ: $\omega \sim N(0,\sigma)$, so that every period a good firm receives a draw from $N(\theta_g,\sigma)$ and a bad firm receives a draw from $N(\theta_b,\sigma)$. In that case, θ^e can take on any value between θ_b and θ_g, because almost every x changes the perceived likelihood of a firm being actually good or bad. Accordingly, a firm's θ^e in period t depends on the entire sequence of x's it received before t. Examining the scarring effect thus would require keeping track of the distribution of the x sequences, and would not be feasible analytically.

This subsection takes a different approach by examining the scarring effect from another angle. In a vintage world with any type of learning, the steady-state proportion of good firms with demand D, denoted as $l_g(D)$, is

$$l_g(D) = \frac{\varphi + \sum_{a=1}^{\bar{a}(D)} \varphi_a(D) h_a(D)}{1 + \sum_{a=1}^{\bar{a}(D)} h_a(D)}. \tag{12}$$

Since φ is exogenous, D affects l_g through three endogenous variables: $\bar{a}$, the maximum firm age; φ_a, the proportion of good firms in a cohort of age a; and h_a, the size of a cohort of age a *relative to the entry size*. Lower demand reduces $\bar{a}$, as some oldest vintages become not viable, and lowers h_a for any a, because more firms have exited as the cohort ages. The impact of demand on φ_a, however, is negative. Lower demand drives stronger selection, and thus raises φ_a for any incumbent cohort.[5] Therefore, D causes two competing effects on l_g through its impact on $\bar{a}$, h_a, and φ_a: on the one hand, increases in φ_a raise l_g; on the other hand, declines in $\bar{a}$ and in h_a reduce l_g by giving incumbent cohorts with higher proportion of good firms less weight in determining l_g.

Does the scarring effect motivated under the all-or-nothing learning capture all these three channels? The $\bar{a}$ channel is fully incorporated: as shown in Fig. 4, firms's maximum life span becomes shorter at a low-demand steady state. However, the h_a channel and the φ_a channel are captured only *partially*. Again, consider Fig. 4: at a low-demand steady state with all-or-nothing learning, h_a is smaller only for cohorts that contain good firms only, and φ_a is higher only for cohorts that possess good firms only when demand is low but some unsure firms when demand is high. By contrast, with more complicated learning, demand impacts h_a and φ_a for *any* age cohorts. Now the question becomes: with more complicated learning, would the scarring effect disappear due to increases in φ_a for more cohorts, or strengthen because of decreases in h_a for more cohorts?

An important remark should be made. Lower demand drives out both potentially good firms and potentially bad firms. It is the exit of potentially bad firms that causes higher φ_a. In that sense, increases in φ_a *add* to the conventional cleansing effect. But, it is the exit of potentially good firms that induces lower h_a: incumbent cohorts become smaller when demand is low, because some potentially good firms that would have remained in operation if demand were high had exited due to low demand—this is the spirit of the scarring effect.

[4] Data are provided by the Bureau of Labor Statistics. The entry and exit series are seasonally adjusted but not employment-weighted. The standard deviations are those of the de-trended variations calculated using the Hodrick–Prescott filter.

[5] With normally distributed noises, θ^e can take on any value between θ_b and θ_g. Each age cohort would feature a cutoff value for θ^e such that firms of this age and with θ^e below this cutoff value choose to exit. Lower demand raises this cutoff θ^e value among each age cohort so that the proportion of good firms is higher for any incumbent cohort.

This point can be emphasized by comparing the following three worlds: one with vintage only, one with learning only, and the other with both vintage and learning. In the world with vintage only as modeled by Caballero and Hammour (1994), recessions create a cleansing effect but no scarring effect. In the world with learning only, recessions bring a cleansing effect by killing off potentially bad firms, and a scarring effect by driving out potentially good firms. In the world with both vintage and learning, the cleansing and scarring effects both carry over, *and* become stronger. Lower $\bar{a}$ destroys marginal vintages in addition to potentially bad firms, which amplifies the cleansing effect, and further reduces older cohorts' weight in determining l_g by driving out the oldest cohorts, which strengthens the scarring effect. Therefore, cleansing and scarring are present as long as learning takes place, with or without vintage.

In summary, the all-or-nothing learning provides a convenient framework to motivate the scarring effect analytically. With more complicated learning, the cleansing and the scarring should carry over and become stronger. Therefore, once again, the question becomes: cleansing and scarring, which effect dominates?

4. With stochastic demand shocks

To evaluate the cleansing and scarring effects quantitatively, this section analyzes a stochastic version of the model, in which demand follows a two-state Markov process with values $[D_h, D_l]$ and a transition probability μ. Accordingly, firms expect the current demand to persist for another period with probability μ, and to change with probability $1-\mu$.

4.1. Calibration

Table 1 summarizes the calibration. With a period as a quarter, β is set to equal 0.99. The value of μ is chosen as 0.95, so that demand switches between a high level and a low level with a constant probability 0.05 per quarter. Bad firms' idiosyncratic productivity, θ_b, is normalized as one. The elasticity of entry cost with respect to entry size, c_1, is chosen based on Goolsbee (1998), who estimates that a 10% increase in demand for equipment raises equipment price by 7.284% (Table VII, p. 143). Accordingly, $c_1 = 0.7284$.

The rest of the parameters are chosen based on data from the U.S. manufacturing sector. In particular, p, φ, γ, and θ_g are calibrated to the observed manufacturing cohort dynamics and productivity differentials. These parameters jointly determine the strengths of learning and creative destruction. With calibrations on p, φ, γ, θ_g, and c_1, changes in demand together with the fixed component of entry cost generate responses in entry and exit, which cause cleansing and scarring. Therefore, D_h, D_l, and c_0 are calibrated to the observed fluctuations in manufacturing plant entry and exit.

4.1.1. Learning pace (p) and prior proportion of good firms (φ)

Assuming that bad firms always exit, (4) implies that the survival rate of a cohort of age a equals $\varphi + (1-\varphi)(1-p)^a$. Dunne et al. (1989) provide corresponding statistics to calibrate p and φ, tracking the exit dynamics of a U.S. manufacturing cohort that entered in 1972. They find that 57.5% of this cohort had exited by 1977 and 78.2% of it had exited by 1982. This imposes two conditions on p and φ:

$$\varphi + (1-\varphi)(1-p)^{19} = 1 - 0.575, \quad \varphi + (1-\varphi)(1-p)^{39} = 1 - 0.782. \tag{13}$$

This gives $p = 0.0538$ and $\varphi = 0.1157$.

4.1.2. Technological pace (γ)

Since only entrants adopt the leading technology in the model, γ is chosen to match the observed growth in entrants' productivity. Jensen et al. (2001) estimate that, after controlling for industry and time effects, the U.S. manufacturing

Table 1
Calibration.

Parameters	value
Quarterly discount factor: β	0.9900
Persistence rate of demand: μ	0.9500
Prior probability of being a good firm: φ	0.1157
Quarterly pace of learning: p	0.0538
Quarterly technological pace: γ	0.0040
Productivity of bad firms: θ_b	1
Productivity of good firms: θ_g	1.7500
Entry cost parameter: c_1	0.7284
Entry cost parameter: c_0	0.1587
High demand: D_h	108.7294
Low demand: D_l	103.9819

entrants' productivity grow by 46.8% from 1963 to 1992 (Table 1, p. 327).[6] The 46.8% increase in entrants' productivity over a 29-year horizon suggests a quarterly technological pace of 0.004.

4.1.3. Idiosyncratic productivity differential (θ_g)

With θ_b normalized as one, θ_g is calibrated to the observed cohort productivity differentials. In the model, productivity differ across birth cohorts due to two factors: the vintage effect, by which younger cohorts have better technology, and the learning of unobservable fixed idiosyncratic productivity, by which older cohorts possess higher proportion of good firms. The latter is defined by Davis and Haltiwanger (1992) as "passive learning". In reality, however, older cohorts may have additional productivity advantages over younger cohorts due to managers' accumulating experiences, workers' learning by doing, technology retooling, and the achieving of economies of scale. These additional effects are defined by Davis and Haltiwanger (1992) as "active learning". Since θ_g captures passive learning only, a careful calibration of θ_g requires controlling for both the vintage effect and the active learning effect.

Accordingly, θ_g is calibrated based on Jensen et al. (2001), who estimate that, after controlling for industry effects, in 1992 the manufacturing incumbents that entered back in 1967 is 15.3% more productive than the new entrants (Table 4, p. 331), although the 1967 vintage is 60.4% *less* productive than the 1992 vintage (Table 2, p. 329). This implies a productivity differential of 75.7% in total to be explained by active learning and passive learning together. Furthermore, Jensen et al. (2001) report that those incumbents that entered in 1967 and have survived through 1992 have grown 14.8% more productive over the 25 years (Table 3, p. 330). This 14.8% productivity growth must be driven by active learning because, under passive learning, plant productivity stays constant. This suggests that active learning causes a productivity differential of 14.8% between new entrants and those incumbents of 25 years old, and leaves a productivity differential of 60.9% to be accounted for by passive learning. Applying these statistics to all-or-nothing learning gives

$$\frac{(\varphi\theta_g + (1-\varphi)(1-p)^{100}\theta_b)}{(\varphi\theta_g + (1-\varphi)\theta_b)(\varphi + (1-\varphi)(1-p)^{100})} = 1.609. \tag{14}$$

Combined with calibrations on p and φ, (14) suggests $\theta_g = 1.75$, a 75% differential between bad and good idiosyncratic productivity.[7]

4.1.4. High demand (D_h), low demand (D_l), and entry cost (c_0)

The values of D_h, D_l, and c_0 are calibrated using the steady-state conditions. Our numerical simulations suggest that, along any sample path with unchanging demand, the dynamics of the model eventually converge to constant entry and exit. The firm distribution at these stable points are similar to those at the steady states, which allows for using the steady-state conditions as approximation.

Let $\overline{a_g}h$, $\overline{a_u}h$ and fh to represent good firms' exit age, unsure firms' exit age, and entry size corresponding to D_h; and let $\overline{a_g}l$, $\overline{a_u}l$, and fl to be those corresponding to D_l. Applying the calibrations on p, φ, γ, and θ_g to (10) gives $\overline{a_g}h - \overline{a_u}h = \overline{a_g}l - \overline{a_u}l = 120$. This leaves $\overline{a_g}h$, $\overline{a_g}l$, fh, and fl to be determined.

The values of $\overline{a_g}h$, $\overline{a_g}l$, fh, and fl are chosen to match the BED statistics on manufacturing plant entry and exit. Fig. 5 plots, in the top two panels, the quarterly entry and exit rates for the U.S. manufacturing sector from 1992 to 2007: the entry rate averages 3.11%, and the exit rate averages 3.45%. Since both series display a declining trend that is not incorporated in the model, they are de-trended using the Hodrick–Prescott filter. The de-trended variations are presented in the bottom two panels of Fig. 5: the de-trended entry rate fluctuates from −0.29% to 0.37%, and the de-trended exit rate varies from −0.36% to 0.34%. This puts the following restrictions on the values of $\overline{a_g}h$, $\overline{a_g}l$, fh, and fl.

First, their implied long-run entry rate and exit rate have to be around 3.11% and 3.45%, respectively. Second, they must match the peak in exit rate and the trough in entry rate at the onset of a recession. That is, when a negative demand shock hits a high-demand equilibrium, the exit rate should rise to 3.79% (3.45% + 0.34%), and the entry rate should drop to 2.82% (3.11% − 0.29%). Third, they must match the trough in exit rate and the peak in entry rate during recovery. Namely, when a positive demand shock hits a low-demand equilibrium, the exit rate should drop to 3.09% (3.45% − 0.36%) and the entry rate should rise to 3.48% (3.11% + 0.37%).

Using a search algorithm that incorporates the related transitory dynamics, we find that these conditions are satisfied for the following combination of parameter values: $\overline{a_g}h = 165$, $\overline{a_g}l = 164$, $fh = 2.6384$, and $fl = 2.5452$. Details are presented in Appendix A. Because $\overline{a_g}h$ and $\overline{a_g}l$ also represent the expected maximum life spans corresponding to a high demand and a low demand, their values are used to calculate the expected values of entry during expansions and during recessions. Applying the values of entry, the entry sizes, and the calibration on c_1 to (7) gives $c_0 = 0.1587$. Applying the calibrations on $\overline{a_g}h$, $\overline{a_g}h$, fh, and fl to the steady-state competitive pricing condition gives $D_h = 108.7294$ and $D_l = 103.9819$.

[6] Jensen et al. (2001) report growths in entrants' productivity in different time periods, among which the 1663–1992 period is the longest. We take statistics over the longest time period reported in Jensen et al. (2001) to calibrate γ, considering that technological pace can vary over time.

[7] This calibration is lower than that by Davis et al. (1999), who assume a high-to-low idiosyncratic productivity ratio of 2.4 based on Bertelsman and Doms (2000) without controlling for the vintage effect or the active learning effect. However, our calibration of θ_g is still consistent with their calibration because, in our model, θ_g is supposed to capture only the effect of passive learning as one of the many effects driving the observed productivity differentials.

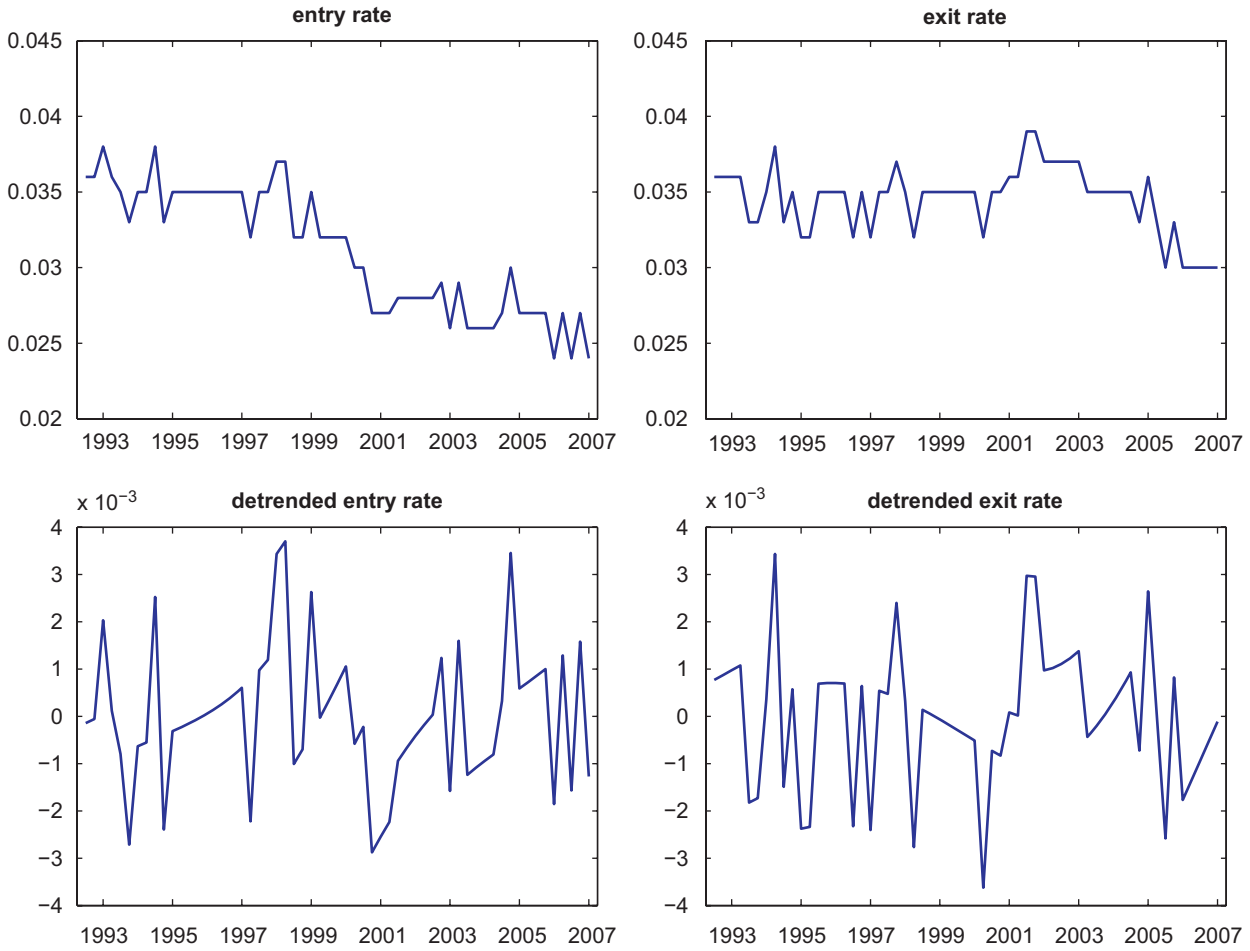

Fig. 5. The quarterly U.S. manufacturing entry and exit rates from 1992 to 2007. The top two panels display the seasonally adjusted, non-employment-weighted raw series; the bottom two panels present the de-trended variations using the Hodrick–Prescott filter. Data source: the business employment dynamics provided by the Bureau of Labor Statistics.

With demand equal to industry total revenue, this implies a revenue differential of 4.57% between expansions and recessions in the U.S. manufacturing sector from 1992 to 2007.

4.1.5. Discussions: average firm age and manufacturing revenue differential

Two remarks should be made on the calibration. First, the calibrated maximum age of good firms equals 165 quarters. Since the model assumes that a firm's vintage stays constant throughout its life span, one may argue that it is hard to believe that some firms would employ the same technology for over 40 years. In reality, new technology adoption takes place not only by entry and exit, but also by incumbents' retooling. The latter, although not incorporated in the model directly, is controlled for when calibrating θ_g, because technology retooling is one of the factors that drive the "active learning" effect on incumbents' productivity growth. In that sense, the model focuses on the technology adoption and the learning of idiosyncratic productivity associated specifically with entry and exit. To further check if the calibration on $\overline{a_g}h$ is reasonable, we compare its implied average firm age with data. Applying $\overline{a_g}h = 165$ to the steady-state firm distribution gives an average firm age of 51 quarters. This is close to the average plant age reported by Faberman (2003), who studies the unemployment records from five states of the U.S. and finds that the sample manufacturing plants age 58 quarters on average.

Second, the calibrations on high demand and low demand based on the observed fluctuations in manufacturing entry and exit suggest a differential of 4.57% in total manufacturing revenue from 1992 to 2007. To check if this calibration is plausible, we examine the 1992–2007 quarterly series of total manufacturing value of shipments at the Census Bureau, and find that the total manufacturing value of shipments fluctuates by about 6% around the trend.[8] The difference between 4.57% and 6%, although quite small, points to the possibility that our calibration underestimates the actual fluctuations in

[8] The examined series are seasonally adjusted, and are de-trended using the Hodrick–Prescott filter.

total manufacturing revenue. This difference can arise from a couple of elements missing from the model but present in reality. For example, demand shocks serve as the only cause for industry fluctuations in the model, while technology shocks are an additional driving force for business cycles in reality. A positive technology shock would drive *further* increases in industry total revenue by raising industry output. Moreover, the wage rate is fixed at one in the model, but tends to comove positively with demand in reality due to, for example, an up-ward slopping labor supply curve (Swanson, 2007). With varying wage rate, output price would have to adjust by more to generate changes in the profit margin that match the observed cyclical entry and exit; consequently, industry total revenue fluctuates by larger magnitude.

4.2. Response to a negative demand shock

With all the parameter values assigned, firm value functions are approximated to simulate the model's responses to stochastic demand shocks. The key computational task is to map F, the firm distribution across ages and idiosyncratic productivity, given demand level D, into a set of value functions $V(\theta^e, a; F, D)$. Unfortunately, F is a high-dimensional object, and it is well known that the numerical solution of dynamic programming problems becomes increasingly difficult as the size of the state space increases. Following Krusell and Smith (1998), our computational strategy shrinks F into a limited set of variables and shows that these variables' laws of motion can approximate the equilibrium behavior of firms in the simulated time series. Details are presented in Appendix A. The value functions and decision rules approximated using these variables enable the investigation of the model's dynamics along any particular path of demand realizations and the study of the model's quantitative implications.

4.2.1. Scarring and cleansing

To assess the effect of a negative demand shock on key variables of the model, the simulation starts with a random firm distribution, and then uses the approximated value functions and decision rules to generate the model's response to the following sequence of demand realizations. Demand stays at D_h until the key variables converge; then, it drops to D_l and persists afterward.

Panel 1 of Fig. 6 illustrates the simulated dynamics of the exit and entry rates in response to a negative demand shock. The quarter labeled 0 denotes the onset of a recession. Panel 1 shows that, when a negative demand shock hits a high-demand long-run equilibrium, the exit rate jumps up, declines afterward, and converges to a level *above* its initial value when demand was high. Put intuitively, a negative demand shock clears out some firms that would stay in operation if demand had remained high. In contrast with the exit rate, the entry rate drops initially, recovers gradually, and converges to a level below its initial value. Panel 1 of Fig. 6 implies that the conventional cleansing effect carries over with an unexpected persistent negative demand shock.

According to the comparative static exercises in Section 3, recessions bring an additional scarring effect that takes place both instantaneously and gradually by worsening the industry composition of idiosyncratic productivity. Panel 2 of Fig. 6 presents the dynamics of the fraction of good firms (l_g) when a negative demand shock hits a high-demand long-run equilibrium. At the onset of a recession, l_g drops due to the "instantaneous" scarring effect. As the recession persists, l_g recovers temporarily, drops again later, and converges eventually to a level below its initial value when demand was high, as suggested by the "lasting scarring" effect. Panel 2 of Fig. 6 implies that the scarring effect also carries over with an unexpected persistent negative demand shock.

Interestingly, the simulated responses of the exit rate, the entry rate, and the fraction of good firms in Fig. 6 all display certain transitory dynamics that have not been captured by the comparative static exercises. The exit rate drops after the initial jump; the entry rate recovers after the initial drop; and the response of l_g appears hump-shaped. These transitory dynamics are driven by the following movements of the exit margins. At the onset of a recession, the exit margins *over shift* to ages younger than the exit ages at the low-demand long-run equilibrium. This is because some old good firms (shown in Fig. 4 as the shaded area) choose to stay at this point by knowing they are good. Their operation raises industry output and lowers output price, causing the exit margins to over shift and the entry size to over drop. As the recession persists, the over-shifted exit margins move back to their stable points quarter by quarter. As unsure firms' exit margin moves to older ages, more good firms are allowed to reach their potential. As good firms' exit margin shifts to older ages, no old good firms exit for several quarters. This gives rise to a temporary "plastic surgery" effect that partially erases the instantaneous scar and drives l_g to rise after its initial drop. Once the exit margins reach their stable points, old good firms and potentially good firms start exiting. At this point, l_g falls again, and converges to a lower level when the industry reaches the low-demand long-run equilibrium.

To summarize, Panels 1 and 2 of Fig. 6 suggest that, despite some transitory dynamics, both the conventional cleansing effect established in Proposition 2, and the scarring effect established in Proposition 3, carry over with an unexpected persistent negative demand shock.

4.2.2. Implications for productivity

With firm-level productivity equal to $A\theta(1+\gamma)^{-a}$, industry average productivity is affected by two components: the leading technology (A), and the firm distribution across vintages (a) and idiosyncratic productivity (θ). Technological

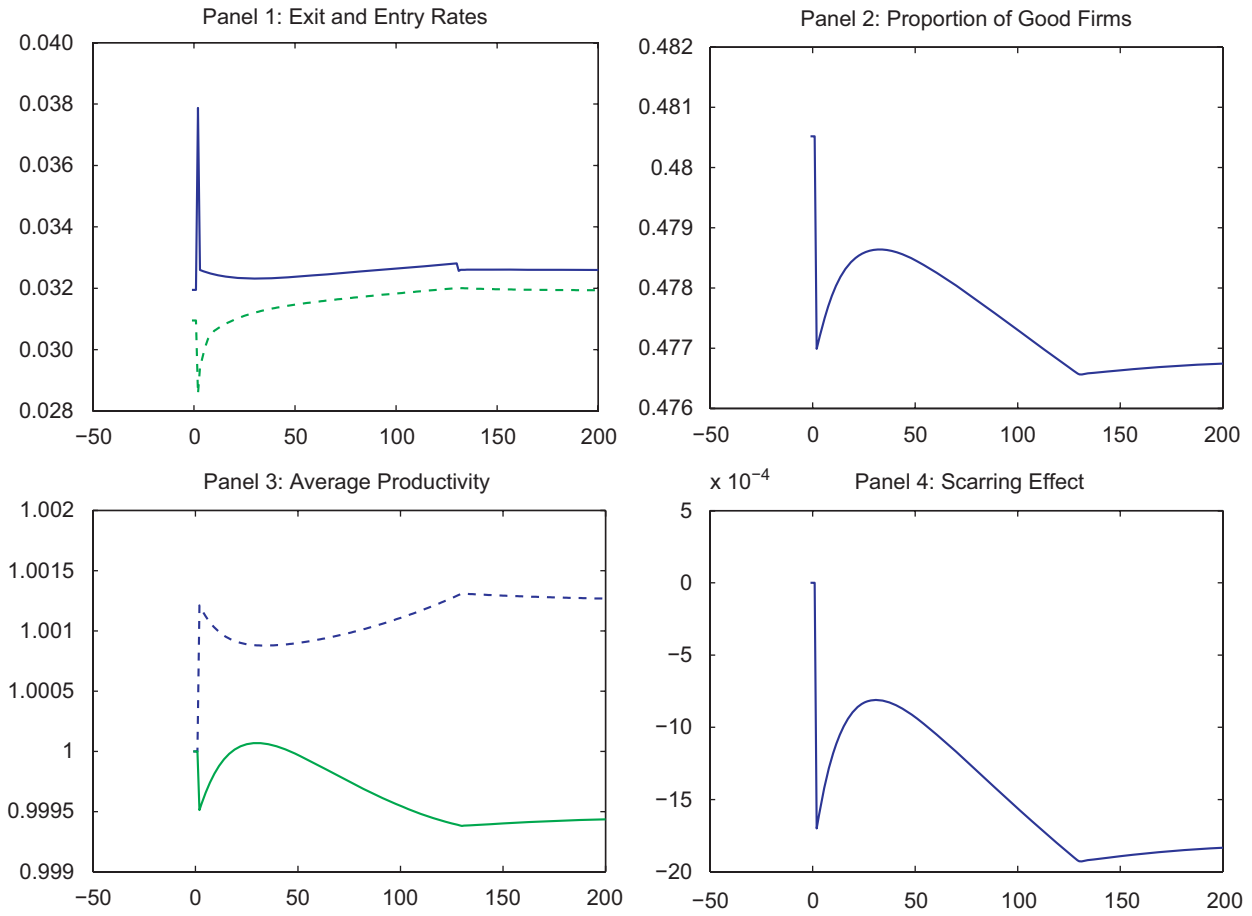

Fig. 6. Simulated responses to a negative demand shock: the horizontal axis denotes quarters, with the quarter labeled 0 as the onset of a recession. In Panel 1, the solid line represents the dynamics of the exit rate, and the dashed line is the dynamics of the entry rate. In Panel 3, the dashed line represents the productivity dynamics driven by the cleansing effect alone, and the solid line denotes the productivity dynamics driven by both the cleansing and scarring effects. Panel 4 plots the distance between the two series plotted in Panel 3. See text for more details.

progress drives A and thus average productivity to grow at rate γ. Demand shocks add fluctuations around this trend by affecting the firm distribution across a and θ.

To analyze the cyclical component of the average productivity, this subsection examines the de-trended average productivity as the average of $\theta(1+\gamma)^{-a}$ over heterogeneous firms. In evaluating this measure, recall that there are two competing effects. On the one hand, the cleansing effect lowers the average a, driving average productivity to rise. On the other hand, the scarring effect reduces the average θ, causing average productivity to fall. To separate these two competing effects, two indexes are created: the average of $\theta(1+\gamma)^{-a}$ over heterogenous firms, denoted as *prod*; and the average of $(1+\gamma)^{-a}$ over heterogenous vintages, denoted as *vin*. Let $f(\theta^e, a)$ to represent the number of firms of age a and with θ^e, *prod* and *vin* are

$$prod = \frac{\sum_f \left(\frac{\theta^e}{(1+\gamma)^a}\right) f(\theta^e, a)}{\sum_f f(\theta^e, a)}, \quad vin = \frac{\sum_f \left(\frac{1}{(1+\gamma)^a}\right) f(\theta^e, a)}{\sum_f f(\theta^e, a)}. \tag{15}$$

Apparently, *prod* is affected by both the cleansing and the scarring effect, while *vin* is driven by the cleansing effect alone. Thus, $prod - vin$ measures the scarring effect on average productivity.

Panel 3 of Fig. 6 traces the percentage change in *prod* and in *vin* when an unexpected persistent negative demand shock hits a high-demand long-run equilibrium. The initial levels of *prod* and *vin* are normalized as one. Panel 3 shows that, at the onset of a recession, *vin* rises to 1.0012, implying that the cleansing effect alone *raises* average productivity by 0.12%; however, *prod* drops to 0.9995, suggesting that the cleansing effect and the instantaneous scarring effect together *lowers* average productivity by 0.05%. As the recession persists, *prod* recovers temporarily due to the "plastic surgery" effect, and declines again as the lasting scarring effect takes place. Eventually, *vin* converges to 1.0013, a 0.13% increase in long-run

average productivity by the cleansing effect alone; but *prod* converges to 0.9994, a 0.06% decline in long-run average productivity under both the cleansing and scarring effects.

Panel 4 of Fig. 6 presents the corresponding dynamics of the scarring effect ($prod - vin$) on average productivity. The scarring effect reduces average productivity by 0.17% at the onset of a recession and by 0.19% in the long run. In summary, Panels 3 and 4 of Fig. 6 suggest that, with plausible calibrations, the scarring effect dominates the cleansing effect and contributes to lower average productivity during recessions.

5. Conclusion

How do recessions affect resource allocation? This paper posits that learning has important consequences for this question. Recessions create, in addition to the conventional cleansing effect, a scarring effect by interrupting businesses' learning of their unobservable idiosyncratic productivity. The scarring effect is evaluated quantitatively based on statistics on entry, exit, and productivity differentials from the U.S. manufacturing sector. A plausible calibration of the model suggests that the scarring effect dominates the cleansing effect, and gives rise to lower average productivity during recessions.

Previous authors have also critiqued the conventional cleansing hypothesis. But the scarring effect differs from their proposed adverse-cleansing effects in important ways. The focus of Ramey and Watson (1997) and Caballero and Hammour (2005) is whether cyclical reallocation is socially efficient: in their models, recessions still promote more productive allocation of resources although associated with lower welfare. The sullying effect proposed by Barlevy (2002) arises from reduced entry rather than concentrated exit. Barlevy (2003) analyzes credit market imperfections rather than the learning of unobservable qualities; moreover, the scarring effect impacts resource allocation both in present times and in the future—a dynamic effect that is missing in Barlevy (2003). Nevertheless, these various adverse-cleansing effects should be viewed as complementary effects that likely amplify each other in reality. For example, during recessions, the credit market frictions can further tighten young businesses' borrowing constraints, so that more potentially good businesses are driven out before they learn; as a result, credit market frictions *deepen* the scarring effect.

A couple of extensions can be added to the model. Firm size can be introduced, allowing firms with better vintages or higher expected idiosyncratic to hire more workers. This modification will generate interesting new predictions. With good firms bigger than unsure firms, a firm would increase its employment when it learns that its idiosyncratic productivity to be good, giving rise to an additional job creation margin driven by learning. In that case, recessions would reduce *later* job creation by driving out potentially good firms at present times. This prediction is consistent with the argument by Caballero and Hammour (2005) that recessions in the U.S. manufacturing sector are usually followed by sluggish job creation during the recovery phase.

The model can also be extended into a general-equilibrium framework. As discussed in Section 2, the exogenous demand shocks can be modeled as arising from consumers' taste shocks on an industry's production goods, or as driven by productivity shocks of down-stream industries that demand an industry's output as one of their inputs. Extension of this model into a general-equilibrium framework will raise interesting new questions. For example, can taste shocks or productivity shocks of plausible sizes generate the observed fluctuations in manufacturing plant entry and exit? Moreover, what is the welfare loss associated with the scarring effect? Such questions are left for future research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jmoneco.2008.12.014.

References

Baden-Fuller, C., 1989. Exits from declining industries and the case of steel castings. Economic Journal 99, 949–961.
Barlevy, G., 2002. The sullying effect of recessions. Review of Economic Studies 69 (1), 65–96.
Barlevy, G., 2003. Credit market frictions and the allocation of resources over the business cycle. Journal of Monetary Economics 50, 1795–1818.
Bartelsman, E.J., Doms, M., 2000. Understanding productivity: lessons from longitudinal microdata. Journal of Economic Literature 38 (3), 569–594.
Bowlus, A., 1995. Matching workers and jobs: cyclical fluctuations in match quality. Journal of Labor Economics 13 (2), 335–350.
Caballero, R., Hammour, M., 1994. The cleansing effect of recessions. American Economic Review 84 (5), 1350–1368.
Caballero, R., Hammour, M., 1996. On the timing and efficiency of creative destruction. Quarterly Journal of Economics 111 (3), 805–852.
Caballero, R., Hammour, M., 2005. The cost of recessions revisited: a reverse-liquidationist view. Review of Economic Studies 72 (2), 313–341.
Caves, R., 1998. Industry organization and new findings on the turnover and mobility of firms. Journal of Economic Literature 36 (4), 1947–1982.
Davis, S., Haltiwanger, J., August 1992. Gross job creation, gross job destruction, and employment reallocation. Quarterly Journal of Economics 7 (3), 819–863.
Davis, S., Haltiwanger, J., 1999. Gross job flows. In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, 41. Elsevier, Amsterdam, pp. 2711–2805.
Davis, S., Haltiwanger, J., Schuh, S., 1996. Job Creation and Destruction. MIT Press, Cambridge.
Dunne, T., Roberts, M.J., Samuelson, L., 1989. The growth and failure of US manufacturing plants. Quarterly Journal of Economics 104 (4), 671–698.
Faberman, R.J., 2003. Job flows and establishment characteristics: variations across metropolitan areas. William Davidson Institute Working Papers 609.
Foster, L., Haltiwanger, J., Syverson, C., 2008. Reallocation, firm turnover, and efficiency: selection on productivity or profitability? American Economic Review 98 (1), 394–425.
Goolsbee, A., 1998. Investment tax incentives, prices, and the supply of capital goods. Quarterly Journal of Economics 113 (1), 121–148.
Hall, R., 2000. Reorganization. Carnegie-Rochester Conference Series on Public Policy 52, 1–22.

Jensen, J.B., McGuckin, R.H., Stiroh, K.J., 2001. The impact of vintage and survival on productivity: evidence from cohorts of US manufacturing plants. Review of Economics and Statistics 83 (2), 323–332.
Jovanovic, B., 1982. Selection and the evolution of industry. Econometrica 50 (3), 649–670.
Krusell, P., Smith Jr., A., 1998. Income and wealth heterogeneity in the macroeconomy. Journal of Political Economy 106 (5), 867–895.
Lucas, R.E., 1978. On the size distribution of business firms. Bell Journal of Economics 9 (2), 508–523.
Mortensen, D., Pissarides, C., 1994. Job creation and job destruction in the theory of unemployment. Review of Economic Studies 61 (3), 397–415.
Pries, M.J., 2004. Persistence of employment fluctuations: a model of recurring firm loss. Review of Economic Studies 71 (1), 193–215.
Pries, M.J., Rogerson, R., 2005. Hiring policies, labor market institutions, and labor market flows. Journal of Political Economy 113 (4), 811–839.
Ramey, G., Watson, J., 1997. Contractual fragility, job destruction, and business cycles. Quarterly Journal of Economics 112 (3), 873–911.
Schumpeter, J.A., 1934. Depressions. In: Douglass, V.B., Edward, C., Seymour, E.H. (Eds.), Economics of the Recovery Program. Whittlesey House, New York, pp. 3–12.
Swanson, E.T., 2007. Real wage cyclicality in the PSID. Scottish Journal of Political Economy 54, 617–647.