

Discussion Paper Series

RIEB

Kobe University

DP2014-37

**Perfect Simulation for Models of
Industry Dynamics**

**Takashi KAMIHIGASHI
John STACHURSKI**

November 21, 2014



Research Institute for Economics and Business Administration

Kobe University

2-1 Rokkodai, Nada, Kobe 657-8501 JAPAN

Perfect Simulation for Models of Industry Dynamics*

Takashi Kamihigashi^{a,b} and John Stachurski^c

November 19, 2014

^a*Research Institute for Economics and Business Administration, Kobe University*

^b*IPAG Business School, Paris, France*

^c*Research School of Economics, Australian National University*

Abstract

In this paper we introduce a technique for perfect simulation from the stationary distribution of a standard model of industry dynamics. The method can be adapted to other, possibly non-monotone, regenerative processes found in industrial organization and other fields of economics. The algorithm we propose is a version of coupling from the past. It is straightforward to implement and exploits the regenerative property of the process in order to achieve rapid coupling.

Keywords: Regeneration, simulation, coupling from the past, perfect sampling

1 Introduction

The dynamics of entry and exit by firms play an essential role in economic theory as well as in real life. Productive new entrants replace unproductive incumbents, rejuvenating overall economic activity. There is a large and growing literature on this economic mechanism (see, e.g., [Hopenhayn and Rogerson \(1993\)](#), [Cooley and Quadrini](#)

*We are grateful for financial support from ARC Discovery Award DP120100321 and the Japan Society for the Promotion of Science. Email: tkamihig@rieb.kobe-u.ac.jp, john.stachurski@anu.edu.au.

(2001) or Melitz (2003)), and much of this literature builds upon the model of entry and exit studied by Hopenhayn (1992). The stationary distribution of entry-exit models of the type studied by Hopenhayn represents a cross-sectional distribution of firms that is both consistent with the definition of equilibrium at any point in time and also invariant over time. For typical parameter values the stationary distribution is uniquely defined but has no analytical solution.¹

Simulation is a useful option for computing the cross-sectional distribution for a given exit policy, since it is not difficult to write down an ergodic Markov process such that its stationary distribution coincides with the cross-sectional stationary distribution. This permits approximate sampling from the stationary distribution simply by running the process from an arbitrary initial value until it is judged to have “nearly” converged. Simulating until the distribution of the state is approximately stationary is referred to as “burn-in.” Unfortunately the length of burn-in required is often the subject of guesswork and heuristics.² Moreover, regardless of how much burn-in is performed, the resulting sample is never exactly stationary, and the size of the error is once again unknown.³

In this paper we show that these problems can be overcome. By using a variation of coupling from the past (CFTP) technique originally due to Propp and Wilson (1996), we show that it is possible to perform perfect sampling—that is, to sample exactly from the stationary distribution of this class of models—for any specified exit threshold. In particular, we develop an efficient algorithm that generates exact, IID draws from the stationary distribution. For each random seed, the algorithm terminates as soon as an exact draw has been generated, and it is guaranteed to terminate in finite time with probability one. Hence there is no need for the heuristics used to judge burn-in time. Moreover, by repeating the algorithm with independent seeds it becomes possible to generate multiple independent draws from the stationary distribution.

Our work draws on a large literature on CFTP that mainly exists outside of economics,

¹The difficulty of obtaining analytical solutions for this model is related to the existence of a positive threshold for productivity at which firms choose to exit. The threshold introduces a nonlinearity that essentially rules out analytical methods.

²While some methods for computing error bounds exist, they are rarely used for two reasons. First, they are nontrivial to compute. Second, these bounds are often highly pessimistic, since any such bounds must address the worst case scenario admitted by the model specification.

³A related issue is that, for a given method, the size of the error is likely to vary with the parameters, since the parameters change the structure of the problem. If the burn-in is not varied accordingly, this is likely to cause bias.

where the technique is often used for models with large but discrete state spaces.⁴ More recently, researchers have developed techniques for implementing CFTP methods in continuous state settings. [Murdoch and Green \(1998\)](#) showed that CFTP can in principle be used in continuous state settings when the underlying Markov process satisfies Doeblin's condition. This condition requires the existence of a nonnegative lower bound function that (a) integrates to a positive value, (b) depends only on the next state, and (c) is pointwise dominated by the transition density function (which depends on both the current state and the next). Theoretical work along the same lines can be found in [Foss and Tweedie \(1998\)](#) and [Athreya and Stenflo \(2003\)](#).

Although these results are fundamental, they can be difficult to apply. [Murdoch and Green \(1998\)](#) admit that their basic method, which is in principle applicable to our model, may have "a limited range of application for two reasons." First, the function associated with Doeblin's condition "may be too small for practical use" to generate exact draws in a reasonable length of time. Second, their method requires the user to draw from scalar multiples of the lower bound transition density and a residual kernel. It can be nontrivial or even impossible to explicitly calculate and draw from these distributions. If approximations are required, this to some degree defeats the purpose of CFTP.

For these reasons, CFTP methods tend to be popular only in specific settings, perhaps the most notable of which is where the underlying Markov process is stochastically monotone. For such processes, efficient and straightforward CFTP methods are available. [Corcoran and Tweedie \(2001\)](#) developed general results on CFTP particularly suitable for monotone Markov processes. An application to economics can be found in [Nishimura and Stachurski \(2010\)](#), where monotonicity makes the algorithm straightforward to implement and analyze.

Here we develop a CFTP algorithm that is designed to produce exact draws from the stationary distribution of Hopenhayn's entry-exit model, which is not monotone. We do however exploit some monotonicity properties from Hopenhayn's model in our algorithm. We show that the algorithm terminates successfully in finite time with probability one by using both the monotonicity of productivity for incumbents and the regenerative property introduced by new entrants. Our algorithm is distinct from Murdoch and Green's method discussed above ([Murdoch and Green \(1998\)](#)), in that it does not use Doeblin's condition, and does not require explicit knowledge of the

⁴Applications range from statistical mechanics to page ranking and the design of peer-to-peer file sharing systems. See, for example, [Propp and Wilson \(1996\)](#), [Kijima and Matsui \(2006\)](#), [Huber \(2003\)](#) and [Levin et al. \(2009\)](#).

transition density.⁵ As long as one can simulate the overall Markov process, one can sample exactly from the stationary distribution using the algorithm.⁶

Aside from related models of industry dynamics, our techniques can also potentially be applied to other non-monotone regenerative models, such as those found in various intertemporal decision problems. One example is the problem of optimal replacement of a part or machine, the performance of which degrades stochastically over time (see, e.g., [Rust \(1987\)](#)).

2 Preliminaries

2.1 The Entry-Exit Model

In this section we briefly review a benchmark model of firm dynamics due to [Hopenhayn \(1992\)](#). The model is set in a competitive industry where entry and exit is endogenously determined. In the model there is a large number of firms that produce a homogeneous good. The firms face idiosyncratic productivity shocks that follow a Markov process on $S := [0, 1]$. The conditional cumulative distribution function for the shock process is denoted by $F(\phi' | \phi)$. Following [Hopenhayn \(1992\)](#), we impose the following restrictions:

Assumption 2.1. F is decreasing in its second argument and, for any $\epsilon > 0$ and any $\phi \in S$, there exists an integer n such that $F^n(\epsilon | \phi) > 0$.⁷

We let P denote the stochastic kernel on $[0, 1]$ corresponding to F . That is, $P(\phi, A) := \int \mathbb{1}_A(\phi')F(d\phi' | \phi)$ for $\phi \in S$ and $A \in \mathcal{B}$, where \mathcal{B} represents the Borel sets on $[0, 1]$ and $\mathbb{1}_A$ is the indicator function of A . Incumbent firms exit the industry whenever their current productivity falls below a reservation value x_t . Letting M_t be the mass of entrants at time t and ν be the Borel probability measure from which the productivity of entrants is drawn, the sequence of firm distributions $\{\mu_t\}$ on S satisfies $\mu_{t+1}(A) = \int P(\phi, A)\mathbb{1}\{\phi \geq x_t\}\mu_t(d\phi) + M_{t+1}\nu(A)$ for all $A \in \mathcal{B}$. At the stationary equilibrium, both x and M are constant, and a stationary distribution μ is a Borel probability⁸ mea-

⁵The assumptions used to show the probability one termination of the algorithm in fact imply Doeblin's condition for some n -step transition, but our proof of this property does not use the latter.

⁶Of course in computer implementations exactness is modulo the errors associated with floating point arithmetic and imperfect random number generators.

⁷ $F^n(\cdot | \phi)$ is the conditional distribution for productivity after n periods, given current productivity ϕ .

⁸We focus only on normalized measures, since other cases are just scalar multiples.

sure μ satisfying

$$\mu(A) = \int P(\phi, A) \mathbb{1}\{\phi \geq x\} \mu(d\phi) + M\nu(A) \quad (A \in \mathcal{B}). \quad (1)$$

It follows from (1) and $\mu(S) = P(\phi, S) = \nu(S) = 1$ that $M = M(x, \mu) := \mu\{\phi \in S : \phi < x\}$. As a result, we can also write (1) as

$$\mu(A) = \int Q(\phi, A) \mu(d\phi) \quad (2)$$

where

$$Q(\phi, A) := P(\phi, A) \mathbb{1}\{\phi \geq x\} + \nu(A) \mathbb{1}\{\phi < x\}. \quad (3)$$

Equation (2) states that μ is a stationary distribution for the stochastic kernel Q in the usual sense of time invariance. As shown by [Hopenhayn \(1992\)](#), the kernel Q has only one stationary distribution. For the purposes of this paper we will treat x as given. For typical parameter values the stationary distribution has no analytical solution.

2.2 Simulation

It is not difficult to produce an ergodic Markov process suitable for simulation such that its stationary distribution (i.e., time-invariant distribution) coincides with the cross-sectional distribution μ in (2). In essence, we need a method for sampling from the stochastic kernel Q . The first step is to simulate from the conditional distribution $P(\phi, \cdot) = F(\cdot | \phi)$. In particular, we seek a random variable U and a function g such that $\mathcal{D}(g(\phi, U)) = F(\cdot | \phi)$ for all $\phi \in S$. (Here $\mathcal{D}(X)$ indicates the distribution of random variable X .) This can be achieved via the inverse transform method, where U is uniform on $[0, 1]$ and $g(\phi, u) = F^{-1}(u | \phi)$.⁹ Now consider the process $\{\Phi_t\}$ defined by

$$\Phi_{t+1} = g(\Phi_t, U_{t+1}) \mathbb{1}\{\Phi_t \geq x\} + Z_{t+1} \mathbb{1}\{\Phi_t < x\} \quad (4)$$

where $\{(U_t, Z_t)\}$ is IID with $\mathcal{D}(Z_t) = \nu$ and $\mathcal{D}(U_t) = \text{Uniform}[0, 1]$. In what follows we call (4) the *simulation model*.

Lemma 2.1. *The simulation model is a Markov process with stochastic kernel Q .*

Proof. Fix $A \in \mathcal{B}$. Applying the law of iterated expectations to (4), we have

$$\begin{aligned} \mathbb{P}\{\Phi_{t+1} \in A\} &= \mathbb{E}[\mathbb{E}_t \mathbb{1}\{\Phi_{t+1} \in A\}] \\ &= \mathbb{E}[\mathbb{E}_t \mathbb{1}\{\Phi_{t+1} \in A\} \mathbb{1}\{\Phi_t \geq x\} + \mathbb{E}_t \mathbb{1}\{\Phi_{t+1} \in A\} \mathbb{1}\{\Phi_t < x\}] \\ &= \mathbb{E}[P(\Phi_t, A) \mathbb{1}\{\Phi_t \geq x\} + \nu(A) \mathbb{1}\{\Phi_t < x\}]. \end{aligned}$$

⁹Here $F^{-1}(\cdot | \phi)$ is the generalized inverse of $F(\cdot | \phi)$. That is, $F^{-1}(u | \phi) := \inf\{z : F(z | \phi) \geq u\}$.

Hence if $\mu_t := \mathcal{D}(\Phi_t)$ for all t , then $\mu_{t+1}(A) = \mathbb{E} Q(\Phi_t, A) = \int Q(\phi, A) \mu_t(d\phi)$. In other words, $\{\Phi_t\}$ is Markov with stochastic kernel Q . \square

Under some additional assumptions one can also show that the simulation model is uniformly ergodic. Hence it can be used to estimate μ via ordinary forward simulation, starting from an arbitrary initial condition. As discussed in the introduction, we adopt a different approach, which permits exact sampling.

3 Perfect Simulation

Before turning to our algorithm it can be helpful to recall some features of CFTP in a simpler setting. To this end we first review a version for finite state Markov chains and then move on to industry dynamics.

3.1 Coupling From the Past

To begin, consider an irreducible aperiodic Markov chain $\{X_t\}$ on finite state space S with unique stationary distribution μ . Suppose that the dynamics of the model are given by $X_{t+1} = h(X_t, W_{t+1})$ where $\{W_t\}$ is IID. If we start at any X_0 and simulate $\{X_t\}$ by drawing W_1, W_2, \dots and successively applying the maps $h(\cdot, W_1), h(\cdot, W_2), \dots$, then the distribution of X_t will converge to (but in general never reach) the stationary distribution μ . A second immediate observation is that if $\{X_t\}$ and $\{X'_t\}$ are two chains simulated using the *same sequence* of shocks $\{W_t\}$, and if $X_\tau = X'_\tau$ for some τ , then $X_t = X'_t$ for all $t \geq \tau$. We say that $\{X_t\}$ and $\{X'_t\}$ *couple* at time τ .

Now consider an IID sequence of shocks $\{W_t\}_{t=-\infty}^0$ indexed on $\{\dots, -2, -1, 0\}$, and let $\{X_t\}_{t=-\infty}^0$ be the process generated from these shocks, starting in the infinite past and terminating at $t = 0$. Intuitively, since the burn-in is infinite, the distribution of X_0 will be exactly μ . On the other hand, because the chain $\{X_t\}_{t=-\infty}^0$ is infinite and hence cannot be simulated, it appears that X_0 is not observable. In fact this is not the case. To understand how X_0 can be observed, suppose that we fix $T \in \mathbb{N}$, take the finite subset of shocks $W_{-T+1}, W_{-T+2}, \dots, W_0$ and then, for each point x in the state space S , construct one version of the process $\{X_t(x)\}_{t=-T}^0$ that starts at this point x at time $-T$ and runs up until time zero. All of these processes $\{X_t(x)\}_{t=-T}^0$ are updated using the same shocks $W_{-T+1}, W_{-T+2}, \dots, W_0$. We will call them *tracking processes*. Because there are only finitely many tracking processes and each is of finite length, it is possible to

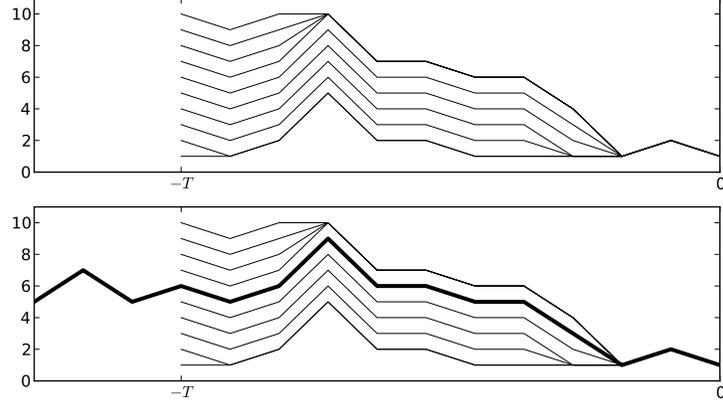


Figure 1: Coalescence of the tracking processes

simulate them. Moreover, following the tracking processes can allow us to pin down the value of X_0 .

To understand the idea, consider first the two different visualizations of the tracking processes shown in the top panel of figures 1 and 2 respectively. Here the state space is $\{1, \dots, 10\}$, and the two figures show different realizations from the same model. Notice that in the simulations, some of the processes couple and then run together up until time zero. In figure 1, all of the tracking processes have coupled by time zero, and only one final value is observed. If this is the case we will say that the tracking processes *coalesce*. In figure 2 multiple paths are still present at time zero, and the processes fail to coalesce.

Now consider again the process $\{X_t\}_{t=-\infty}^0$ generated from the entire sequence of shocks $\{W_t\}_{t=-\infty}^0$. Since this process must pass through one point x in the state space at $-T$, and since it receives the same shocks $W_{-T+1}, W_{-T+2}, \dots, W_0$ as the tracking processes from that time forwards, over the period $-T, -T+1, \dots, 0$ it must follow the same path as the tracking process $\{X_t(x)\}_{t=-T}^0$ that started at x . Of course we do not know which of the tracking processes it is following. However, *if* the tracking processes coalesce, then there is only *one* final observation. This observation must be a realization of the time zero value X_0 of the process $\{X_t\}_{t=-\infty}^0$ that started in the infinite past, and is therefore a draw from the stationary distribution μ . Such an outcome is illustrated in the bottom panel of figure 1. For comparison, an unsuccessful outcome is illustrated in the bottom panel of figure 2. Here there are three final values, and we do not know which is the time zero value of the process $\{X_t\}_{t=-\infty}^0$. In this case we can take one step

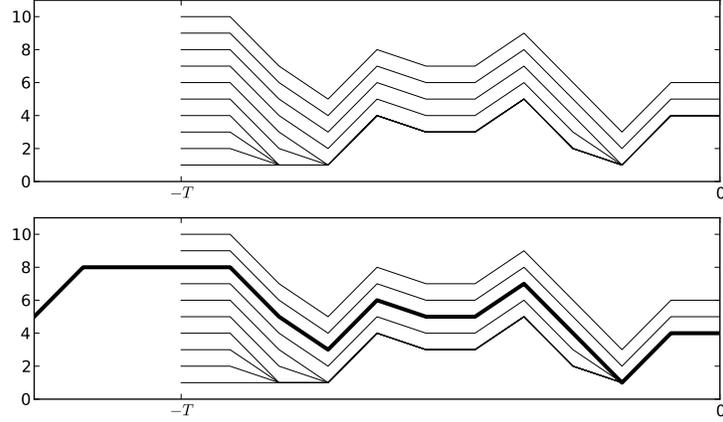


Figure 2: The tracking processes failing to coalesce

further back in time, drawing the additional shock W_{-T} while preserving the existing shocks $W_{-T+1}, W_{-T+2}, \dots, W_0$, recalculate the tracking processes, test for coalescence, and so on. This procedure will eventually terminate with an exact draw from μ .¹⁰

3.2 CFTP with the Simulation Model

While the technique presented in the previous section works when there are a finite number of states, it is clearly not possible to directly implement the same idea when the state space is infinite. However, variations on the idea can potentially be found by exploiting the structure of a given model. In particular, for the simulation model defined in (4), we now show how the monotonicity of g and the renewal component can be combined to pin down the final value X_0 (which is Φ_0 in this case).

To begin, let $\{(U_t, Z_t)\}_{t \in \mathbb{Z}}$ be an infinite sequence of IID shocks indexed on \mathbb{Z} and with each pair (U_t, Z_t) having the product distribution $\text{Uniform}[0, 1] \times \nu$. To simplify notation we will let $g_t := g(\cdot, U_t)$, so that, for example, $g_t \cdots g_1 \phi := g_t \circ g_{t-1} \circ \cdots \circ g_1(\phi)$ is exogenous productivity at t , given time zero productivity $\phi \in S$. To further simplify notation, let

$$h_t(\phi) := g(\phi, U_t) \mathbb{1}\{\phi \geq x\} + Z_t \mathbb{1}\{\phi < x\},$$

so that (4) becomes $\Phi_{t+1} = h_{t+1} \Phi_t := h_{t+1}(\Phi_t)$.

¹⁰Using finiteness of the state space, aperiodicity and irreducibility, it can be shown that termination occurs in finite time with probability one.

Now fix $T \geq 1$. For each $\phi \in S$, there is a corresponding “tracking process” that starts at time $-T$ with value ϕ , and then updates with maps $h_{-T+1}, h_{-T+2}, \dots, h_0$, obtaining the value $h_0 \cdots h_{-T+1} \phi$ at time zero. We say that the tracking processes *coalesce* if, for some $T \in \mathbb{N}$, the set of final values

$$h_0 \cdots h_{-T+1}(S) := \{h_0 \cdots h_{-T+1} \phi : \phi \in S\} \quad (5)$$

is a singleton. What we will now show is that under mild conditions coalescence occurs with probability one, and, moreover, that it is not necessary to keep track of the full continuum of tracking processes in order to find the value of the singleton. In particular, we show that, conditional on a certain event described below, the set of final values $h_0 \cdots h_{-T+1}(S)$ has only *finitely* many possibilities. Hence coalescence occurs whenever these finite possibilities take the same value. All of these finite possibilities are computable. To begin describing them, let $T > 1$ be given, let

$$\Sigma_T := \{k \in \mathbb{N} : 1 \leq k < T \text{ and } g_{-T+k} \cdots g_{-T+2} \cdot g_{-T+1} 1 < x\},$$

and let $\sigma_T := \min \Sigma_T$. Intuitively, σ_T is the number of periods that an incumbent firm survives, given that it starts at time $-T$ with maximal productivity 1 and faces the shock sequence $\{U_t\}_{t \in \mathbb{Z}}$. Clearly σ_T is only defined when Σ_T is nonempty. However, the probability that Σ_T is nonempty converges to one as $T \rightarrow \infty$ by assumption 2.1. Moreover, it is remarkable that if Σ_T is nonempty, then the set $h_0 \cdots h_{-T+1}(S)$, which contains the final values of the tracking processes started at $-T$, can have only finitely many values:

Lemma 3.1. *If Σ_T is nonempty, then $h_0 \cdots h_{-T+1}(S) \subset \Lambda_T$, where the latter is the finite set given by*

$$\Lambda_T := \{h_0 \cdots h_{-T+k+1} Z_{-T+k} : k = 1, \dots, \sigma_T + 1\}. \quad (6)$$

The proof of lemma 3.1 is given in section 4. The intuition for the result can be obtained by considering figure 3. In the figure, $T = 10$. Tracking processes are plotted for 50 different initial values of $\phi \in S$. (Ideally, tracking processes would be plotted from every $\phi \in S$, but this is clearly impossible.) For this particular realization of shocks, the set Σ_T is nonempty because the process starting from 1 at time -10 falls below x at $t = -2$ (and hence $\sigma_T = 10 - 2 = 8$). As is clear from the figure, the fact that the process starting from 1 at time -10 falls below x at $t = -2$ implies that *all* tracking processes fall below x at least once between -10 and -2 (recall that the productivity of incumbents is monotone). Moreover, if any collection of tracking processes fall below x at some point in time t , they subsequently couple, taking the common value Z_{t+1} at

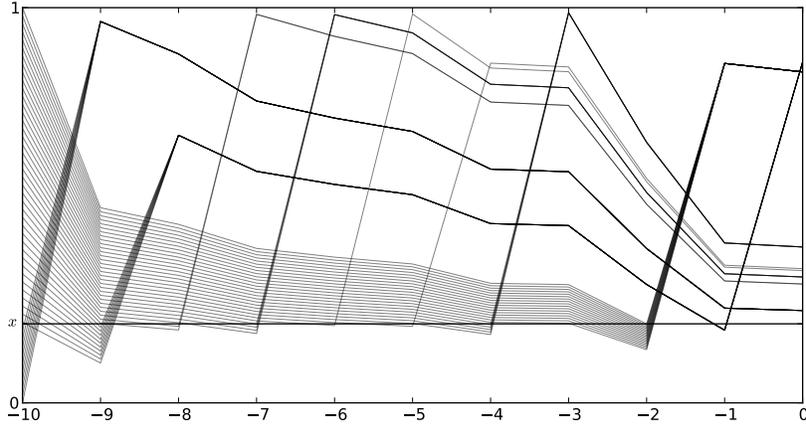


Figure 3: Tracking processes with $T = 10$ and $\sigma_T = 8$

$t + 1$ and being equal from then on. As a result, by $t = -1$, there are at most $\sigma_T + 1 = 9$ distinct tracking processes. Their time zero values are included in the set Λ_T defined in lemma 3.1. In particular, Λ_T is the time zero values of the processes that start below x at dates $-10, -9, \dots, -2$.

To see the importance of lemma 3.1, let $\{\Phi_t\}_{t \in \mathbb{Z}}$ be a stationary, doubly-indexed process on the same probability space as $\{(U_t, Z_t)\}_{t \in \mathbb{Z}}$ that obeys $\Phi_{t+1} = h_{t+1} \Phi_t$ for all $t \in \mathbb{Z}$. The common marginal distribution of Φ_t is μ . Since Φ_{-T} lies somewhere in S , we know that $\Phi_0 = h_0 \cdots h_{-T+1} \Phi_{-T} \in h_0 \cdots h_{-T+1}(S)$. Moreover, if the set Σ_T is nonempty, then lemma 3.1 yields the inclusion $h_0 \cdots h_{-T+1}(S) \subset \Lambda_T$, and Φ_0 lies in the finite observable set Λ_T . In particular, if Λ_T is a singleton, then the value of Φ_0 is revealed as the value of that singleton.

Figures 4 and 5 show simulations with successful and unsuccessful coalescence respectively. In each figure, the top panel shows only the tracking processes. (As with figure 3, the full continuum of tracking processes cannot be plotted, so we show only 50.) The bottom panel shows the tracking processes and the path of $\{\Phi_t\}$. In reality, the path of $\{\Phi_t\}$ is not observed. However, in figure 4, there is only one final, coalesced value, and Φ_0 must take this value. Hence Φ_0 is observed. On the other hand, in figure 5, Φ_0 is equal to one of two final values, and we have no way of identifying which one it is.

Now let us consider how to use our results to sample from μ by generating observations of Φ_0 . In order to avoid conditioning on coalescence by a certain point in time, we wish to reveal the value of Φ_0 for every random seed. This can be done by fixing

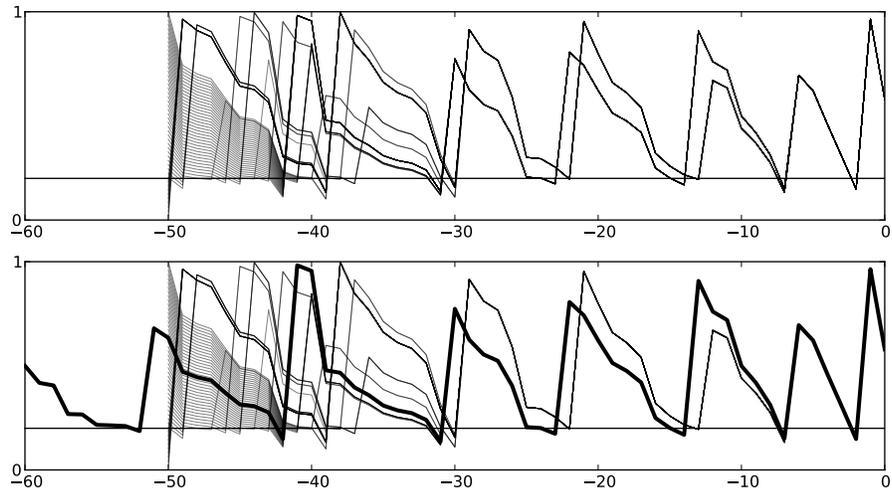


Figure 4: Successful coalescence from $T = 50$

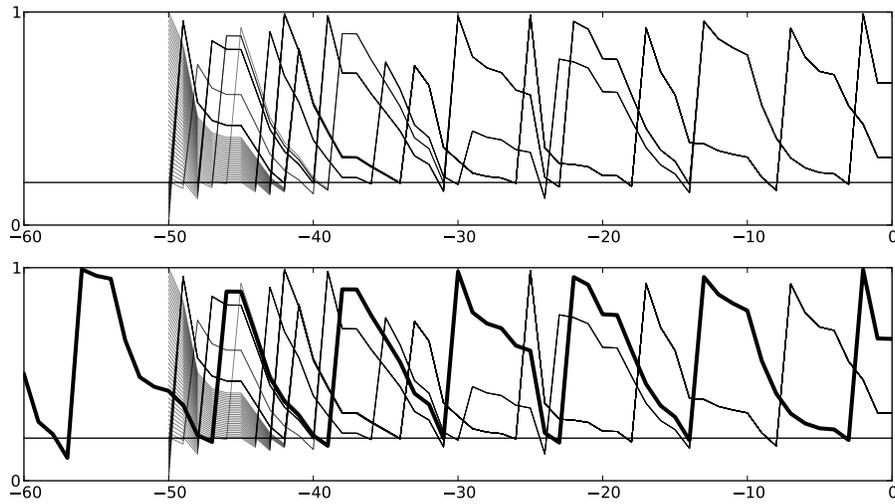


Figure 5: Failure of coalescence

the seed, which determines the values of the shock processes, and then taking T larger and larger until coalescence occurs. Algorithm 1 gives details.

Algorithm 1: Generates an exact draw from μ

```

fix  $T$  to be an integer greater than 1;
draw  $(U_0, Z_0), \dots, (U_{-T+1}, Z_{-T+1})$  independently from their distributions;
repeat
    compute the set  $\Sigma_T$ ;
    if  $\Sigma_T$  is nonempty then
        compute the set  $\Lambda_T$ ;
        if  $\Lambda_T$  is a singleton then
            set  $X$  to be the value of that singleton;
            break;
        end
    end
    draw  $(U_{-T}, Z_{-T})$  and append to list  $(U_0, Z_0), \dots, (U_{-T+1}, Z_{-T+1})$ ;
    set  $T = T + 1$ ;
end
return  $X$ ;

```

We still need to show that (a) the distribution of X is μ , and (b) that the algorithm will terminate with probability one. We now show that these claims are true under the following condition, which is satisfied by many standard distributions.

Assumption 3.1. If $G \subset S$ is a nonempty open set, then $\nu(G) > 0$.

Let $T^* := \min\{T \in \mathbb{N} : \Sigma_T \text{ is nonempty and } \Lambda_T \text{ is a singleton}\}$. In other words, T^* is the length of the simulation run required for algorithm 1 to terminate. Evidently T^* is a random variable.

Proposition 3.1. *If assumption 3.1 holds, then*

1. *There exists a $\gamma \in (0, 1)$ such that $\mathbb{P}\{T^* > t\} \leq \gamma^t$.*
2. *The random variable X returned by algorithm 1 has distribution μ .*

Two comments are in order. First, item 1 tells us in particular that $\mathbb{P}\{T^* < \infty\} = 1$. In other words, the algorithm terminates in finite time with probability one. Thus proposition 3.1 not only gives probability one occurrence but also provides a geometric rate.

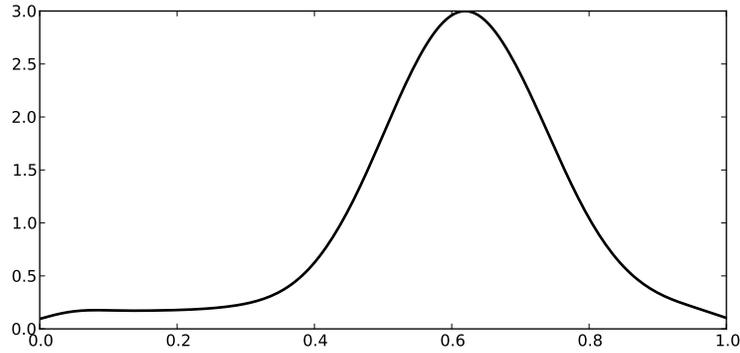


Figure 6: Stationary density

Second, in view of item 2, algorithm 1 gives us a way to generate an exact sample from μ . If we then repeat the process with independent shocks we will generate multiple independent draws. The proof of proposition 3.1 is given in section 4.¹¹

The web page https://github.com/jstac/hh_sampling contains a simple implementation of algorithm 1. We tested the code by following Hopenhayn and Rogerson (1993) in taking the distribution ν for new entrants to be uniform, and the process for incumbents to be $\Phi_{t+1} = g_{t+1}\Phi_t = a + \rho\Phi_t + \epsilon_{t+1}$ where $\{\epsilon_t\}$ is IID with distribution $N(0, \sigma^2)$. To bound the process we added reflecting barriers at 0 and 1. The parameters were set to $a = 0.36$, $\rho = 0.4$ and $\sigma = 0.1$, while x was set to 0.49, so that approximately 40% of incumbents exit within 5 years (Hopenhayn, 1992, p. 1127). For these parameters, running the program on a standard workstation without parallelization produces about 36,000 independent draws from μ per second.¹²

Figure 6 shows the density computed from 36,000 observations combined with a standard nonparametric kernel density estimator (using a Gaussian kernel). Figure 7 shows a 95% confidence set for the cumulative distribution function corresponding to μ , based on the same observations and calculated using the Kolmogorov distribution of the sup norm deviation between true and empirical cdfs. The Kolmogorov result is applicable here because the draws are exact and IID. The true distribution function lies entirely between the two bands with 95% probability.

¹¹A final point of intuition before the proofs is that on first reading it might appear that X is a draw from the distribution ν of new entrants, rather than the stationary distribution μ , since coalescence to a single value always occurs after a re-entry event. However, this is not the case because the algorithm still runs on from this event until we reach time zero, at which point the final value is recorded. See figure 4.

¹²Our workstation has a 2.67GHz Intel CPU and 4 gigabytes of RAM.

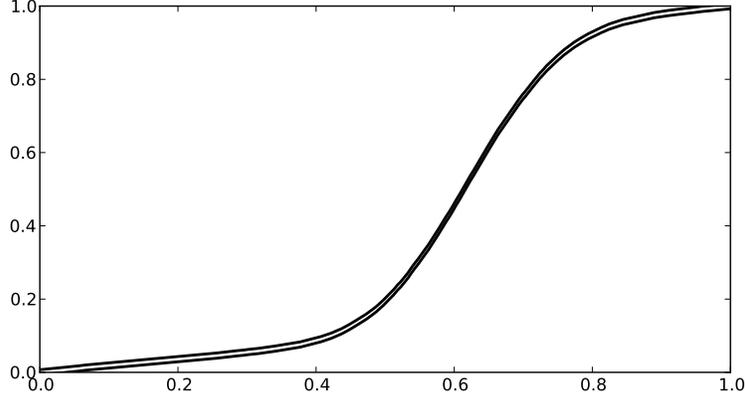


Figure 7: 95% confidence set for the stationary distribution

4 Proofs

In the following proofs T is fixed, and we write σ for σ_T to simplify notation.

Proof of lemma 3.1. Let Σ_T be nonempty. As a first step, we show that if $\phi \geq x$, then there exists a $j \in \{1, \dots, \sigma\}$ such that $h_{-T+j} \cdots h_{-T+1} \phi < x$. To see that this is so, fix $\phi \geq x$ and suppose that the statement fails. In other words, $h_{-T+j} \cdots h_{-T+1} \phi \geq x$ for $j \in \{1, \dots, \sigma\}$. We know that if $y \geq x$, then $h_i y = g_i y$. It follows that $h_{-T+\sigma} \cdots h_{-T+1} \phi = g_{-T+\sigma} \cdots g_{-T+1} \phi$. But then

$$x \leq h_{-T+\sigma} \cdots h_{-T+1} \phi = g_{-T+\sigma} \cdots g_{-T+1} \phi \leq g_{-T+\sigma} \cdots g_{-T+1} 1 < x,$$

where the second inequality is due to monotonicity of g_i , and then third is by the definition of σ . Contradiction.

To complete the proof, pick any $\phi \in S$. Our claim is that $h_0 \cdots h_{-T+1} \phi \in \Lambda_T$. Suppose first that $\phi < x$. In this case we have $h_0 \cdots h_{-T+1} \phi = h_0 \cdots h_{-T+2} Z_{-T+1}$, which is an element of Λ_T . Next, suppose that $\phi \geq x$. In light of the preceding argument, there exists a $j \in \{0, \dots, \sigma\}$ with $h_{-T+j} \cdots h_{-T+1} \phi < x$, and hence

$$h_{-T+j+1} \cdot h_{-T+j} \cdots h_{-T+1} \phi = Z_{-T+j+1},$$

from which we obtain

$$\begin{aligned} h_0 \cdots h_{-T+1} \phi &= h_0 \cdots h_{-T+j+2} \cdot h_{-T+j+1} \cdot h_{-T+j} \cdots h_{-T+1} \phi \\ &= h_0 \cdots h_{-T+j+2} Z_{-T+j+1}. \end{aligned}$$

Since $j \in \{0, \dots, \sigma\}$, the right-hand side is an element of Λ_T . This completes the proof. \square

Proof of proposition 3.1. We start with the first claim, which is existence of a $\gamma \in (0, 1)$ such that $\mathbb{P}\{T^* > t\} \leq \gamma^t$.

Let n be an integer such that $F^n(x | 1) > 0$, existence of which is due to assumption 2.1. Fixing $j \in \mathbb{N}$, let

$$E_j := \{g_{-(j-1)n-1} \cdots g_{-jn} \mathbf{1} < x\} \cap \{Z_{-(j-1)n-1} < x, \dots, Z_{-jn} < x\}.$$

The events $\{g_{-(j-1)n-1} \cdots g_{-jn} \mathbf{1} < x\}$ and $\{Z_{-(j-1)n-1} < x, \dots, Z_{-jn} < x\}$ are independent because the first event depends only on $U_{-(j-1)n-1}, \dots, U_{-jn}$ and the second depends only on $Z_{-(j-1)n-1}, \dots, Z_{-jn}$. As a result,

$$\delta := \mathbb{P}(E_j) = F^n(x | 1) \nu([0, x])^n.$$

The constant δ is strictly positive as a result of assumption 3.1. We claim that if the event E_j occurs, then Σ_{jn+1} is nonempty and Λ_{jn+1} is a singleton. To simplify notation, we treat only the case of $j = 1$.

So suppose that E_1 occurs. Clearly Σ_{n+1} contains n , and hence is nonempty. To see that Λ_{n+1} is a singleton, observe that since $\sigma = \sigma_{n+1}$ is the smallest element of Σ_{n+1} , we must have $\sigma \leq n$. As a consequence,

$$\begin{aligned} \Lambda_{n+1} &= \{h_0 \cdots h_{-n+k} Z_{-n-1+k} : k = 1, \dots, \sigma + 1\} \\ &\subset \{h_0 \cdots h_{-n+k} Z_{-n-1+k} : k = 1, \dots, n + 1\}. \end{aligned}$$

We claim that on the set E_1 we have

$$h_0 \cdots h_{-n+k} Z_{-n+k-1} = Z_0 \quad \text{for any } k \in \{1, \dots, n + 1\}. \quad (7)$$

To prove that (7) holds, observe that on E_1 the values Z_{-1}, \dots, Z_{-n} are all less than x . As a result, we have

$$\begin{aligned} h_0 Z_{-1} &= Z_0 \\ h_0 h_{-1} Z_{-2} &= h_0 Z_{-1} = Z_0 \\ h_0 h_{-1} h_{-2} Z_{-3} &= h_0 h_{-1} Z_{-2} = h_0 Z_{-1} = Z_0 \end{aligned}$$

and so on. Together, these equalities give (7). As a consequence, we conclude that Σ_{n+1} is nonempty and Λ_{n+1} is a singleton whenever E_1 occurs, and, more generally, Σ_{nj+1}

is nonempty and Λ_{nj+1} is a singleton whenever E_j occurs. The events E_1, E_2, \dots are independent and have positive probability δ . Using the definition of T^* , we then have

$$\mathbb{P}\{T^* > nj\} = \mathbb{P}\{T^* \geq nj + 1\} \leq \mathbb{P} \cup_{i=1}^j E_i^c = (1 - \delta)^j$$

for all $k \in \mathbb{N}$. Setting $\gamma := (1 - \delta)^{1/n}$ gives the first result stated in proposition 3.1

The second claim in proposition 3.1 is that the value X returned by algorithm 1 has distribution μ . It is clear from the algorithm that X is the value contained in Λ_T when the latter is a singleton. In the discussion preceding the proposition it was shown that when Λ_T is a singleton the value of the singleton is Φ_0 . Hence $X = \Phi_0$. Moreover, $\{\Phi_t\}$ is by assumption a strictly stationary process, and hence the distribution of Φ_0 is exactly μ . \square

References

- Athreya, Krishna B and Örjan Stenflo (2003) "Perfect sampling for Doeblin chains," *Sankhyā: The Indian Journal of Statistics*, pp. 763–777.
- Cooley, Thomas F and Vincenzo Quadrini (2001) "Financial markets and firm dynamics," *American Economic Review*, pp. 1286–1310.
- Corcoran, JN and RL Tweedie (2001) "Perfect sampling of ergodic Harris chains," *The Annals of Applied Probability*, Vol. 11, pp. 438–451.
- Foss, Sergey G and RL Tweedie (1998) "Perfect simulation and backward coupling," *Stochastic Models*, Vol. 14, pp. 187–203.
- Hopenhayn, Hugo A (1992) "Entry, exit, and firm dynamics in long run equilibrium," *Econometrica*, pp. 1127–1150.
- Hopenhayn, Hugo and Richard Rogerson (1993) "Job turnover and policy evaluation: A general equilibrium analysis," *Journal of Political Economy*, pp. 915–938.
- Huber, Mark (2003) "A bounding chain for Swendsen-Wang," *Random Structures & Algorithms*, Vol. 22, pp. 43–59.
- Kijima, Shuji and Tomomi Matsui (2006) "Polynomial time perfect sampling algorithm for two-rowed contingency tables," *Random Structures & Algorithms*, Vol. 29, pp. 243–256.

- Levin, David A, Yuval Peres, and Elizabeth L Wilmer (2009) "Markov chains and mixing times. With a chapter by James G. Propp and David B. Wilson," *American Mathematical Society, Providence, RI*.
- Melitz, Marc J (2003) "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, Vol. 71, pp. 1695–1725.
- Murdoch, Duncan J and Peter J Green (1998) "Exact sampling from a continuous state space," *Scandinavian Journal of Statistics*, Vol. 25, pp. 483–502.
- Nishimura, Kazuo and John Stachurski (2010) "Perfect simulation of stationary equilibria," *Journal of Economic Dynamics and Control*, Vol. 34, pp. 577–584.
- Propp, James Gary and David Bruce Wilson (1996) "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Random Structures and Algorithms*, Vol. 9, pp. 223–252.
- Rust, John (1987) "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica*, pp. 999–1033.