# 14 Manuscripts on Progress Presentation

## Topic:

Entry/Exit Model, Power law and A conjecture

# 1 Intro to Firm Size Distribution and Power Law*

## 1.1 Preparation

## 1.2 Potential Contexts on slides

## 1.3 Manuscripts

# 2 Intro to Markov Chain-give an example*

## 2.1 Preparation*

### Def. (Stochatic (markov) matrices)

A stochastic matrix (or Markov matrix) is an $n \times n$ square matrix $P = (p_{i,j})$ such that

1. Each element of $P$ is nonnegative ($p_{i,j} \geq 0$), and
2. each row of $P$ sums to 1, i.e., $p_{i,1} + p_{i,2} + \cdots + p_{i,n} = 1$ for $i \in \{1, \ldots, n\}$.

#### Remarks

Each row of $P$, $p_{i,j}(j)$ can be regarded as a probability mass function over $n$ possible outcomes.

### Proposition

If $P$ is a stochastic matrix, then the $k$-th power $P^k$ for all $k \in \mathbb{N}$ is also a stochastic matrix.

### Def. (Markov Chains)

Let $S$ be a finite set with $n$ elements $\{x_1, \ldots, x_n\}$, where the set $S$ is called the **state space** and $x_1, \ldots, x_n$ are the **state values**.

A **Markov chain** $\{X_t\}$ on $S$ is a sequence of random variables on $S$ (from probability sample space $\Omega$) that have the **Markov property**: for any date $t$ and any state $y \in S$,

$$\mathcal{P}\{X_{t+1} = y | X_t\} = \mathcal{P}\{X_{t+1} = y | X_t, X_{t-1}, \dots\} \tag{1}$$

The dynamics of a Markov chain are fully determined by the set of values

$$P(x, y) = \mathcal{P}\{X_{t+1} = y | X_t = x\} \; (x, y \in S) \tag{2}$$

## Example: Job search model

### Set up

1. Consider a worker who, at any given time $t$, is either unemployed (state $0$) or employed (state $1$).
2. Suppose that, over a one month period,
   - An unemployed worker finds a job with probability $\alpha \in (0, 1)$.
   - An employed worker loses her job and becomes unemployed with probability $\beta \in (0, 1)$.

### Markov Chain and Markov matrix

1. Sample space $E = \{unemployed, employed\}$;
2. State space: $S = \{0, 1\}$;
3. Markov Chain: $\{X_t = x\}, \forall x \in S$ and $X_t$ is a random variable $X_t : E \to S$:
   - $X_t(unemployed) = 0$;
   - $X_t(employed) = 1$
4. Transition probabilities: $P(0, 1) = \alpha$ and $P(1, 0) = \beta$.
   - In matrix form, Markov Matrix

$$\begin{aligned} P &= \begin{pmatrix} P(0,0) & P(0,1) \\ P(1,0) & P(1,1) \end{pmatrix} \\ &= \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix} \end{aligned} \tag{3}$$

# 2.2 Potential Contexts on slides

# 2.3 Manuscripts

# 3 Intro to Stationary distribution, Aperiodicity and irreducibility

# 3.1 Preparation

## 3.1.1 Marginal distribution

Suppose that (given)

1. $\{X_t\}$ is a Markov chain with stochastic matrix $P$;
2. the distribution of $X_t$ is known to be $\psi_t$.

Q: what is the distribution of $\{X_{t+1}\}$ or, more generally, of $X_{t+m}$?

### Solution

Let $\psi$ be the distribution of $X_t$ for $t = 0, 1, 2, \ldots$

**Step 1: To find $\psi_{t+1}$, given $\psi_t$ and $P$.**

By the Law of toal probability, we can decompose the probability that $X_{t+1} = y$:

$$P\{X_{t+1} = y\} = \sum_{x \in S} P\{X_{t+1} = y | X_t = x\} \cdot P\{X_t = x\} \tag{4}$$

which means that to get the probability of being at $y$ tomorrow, we account for all ways this can happen and sum their probabilities.

We can rewrite (3) in terms of marginal and conditional probabilities

$$\psi_{t+1}(y) = \sum_{x \in S} P(x, y) \psi_t(x) \tag{5}$$

**Step 2: Represent n equations by the matrix form**

There are $n$ possible values of $y$ in the state space $S$, so one equation $(4)$ for each $y \in S$, there are $n$ such equations.

If we think of $\psi_{t+1}$ and $\psi_t$ as row vectors, these $n$ equations are summuarized by the matrix expression

$$\psi_{t+1} = \psi_t P \tag{6}$$

which means that to move the distribution forward one unit of time, we postmultiply by $P$.

**Step 3:**

By repeatiing this $m$ time, we move forward $m$ steps into the future.

By iterating on $(5)$, we get the expression

$$\psi_{t+m} = \psi_t P^m \tag{7}$$

where $P^m$ is the $m$-th power of $P$.

**Step 4:**

For a special case, if $\psi_0$ is the initial distribution from which $X_0$ is drawn, then $\psi_0 P^m$ is the distribution of $X_m$.

Let's repeat it

$$X_0 \sim \psi_0 \implies X_m \sim \psi_0 P^m \tag{8}$$

and more generally,

$$X_t \sim \psi_t \implies X_{t+m} \sim \psi_t P^m \tag{9}$$

### Multiple step transition probabilities

Since the probability of transitioning from $x$ to $y$ in one step is $P(x, y)$, the probability of transitioning from $x$ to $y$ in $m$ steps is

$$P^m(x, y) \tag{10}$$

Which is the $(x, y)$-th element of the $m$-th power of $P$.

#### Reasons behind

Consider $\psi_t$ with putting all probability on state $x$, that is, we put $1$ in the $x$-th position and $0$ elsewhere.

Inserting this into $(8)$, conditional on $X_t = x$, the distribution of $X_{t+m}$ is the $x$-th row of $P^m$.

In particular, we have

$$P\{X_{t+m} = y\} = P^m\{x, y\} = (x, y) - th\ element\ of\ P^m \tag{11}$$

## 3.1.2 Irreducibility and aperiodicity

They are central concepts of modern Markov chain theory.

### Def. (Two states communicate)

Let $P$ be a fixed stochastic matrix.

Two states $x$ and $y$ are said to communicate with each other if there exist positive integers $j$ and $k$ such that

$$P^j(x, y) > 0\ and\ P^k(y, x) > 0 \tag{4}$$

which means precisely that

1. state $x$ can be reached eventually from state $y$,
2. state $y$ can be reached eventually from state $x$.

### Def. (Irreducibility)

The stochastic matrix $P$ is called **irreducible** if all states communicate, that is, $x$ and $y$ communicate for **all** $(x, y)$ in $S \times S$.

#### Remark

Irreducibility is important in terms of long run outcomes. e.g., poverty is a life sentence in the second graph but not the first.

### Def. (Period)

The period of a state $x$ is the greatest common divisor of the set of integers

$$D(x) = \{j \geq 1 | P^j(x, x) > 0\} \tag{5}$$

### Def. (Aperiodicity: formal)

A stochastic matrix is called aperiodic if the period of every state is 1, and periodic otherwise.

## 3.1.3 Stationary distributions

As seen in $(5)$, we can shift probabilities forward one unit of time via post multiplication by $P$.

Some distributions are invariant under this updating process, and such distributions are called stationary, or invariant.

### Def. (Stationary, Invariant)

A distribution $\psi^*$ on $S$ is called stationary for $P$ if

$$\psi^* = \psi^* P \tag{6}$$

**Remarks**

1. From $(14)$, we can get $\psi^* = \psi^* P^t$ for all $t$.

which tells us an **important fact**:

If the distribution of $X_0$ is a stationary distribution, then $X_t$ will have this same distribution for all $t$.

2. Hence, stationary distributions have a natural interpretation as stochastic steady states.
3. Mathematically, a stationary distribution is a fixed point of $P$ when $P$ is thought of as the map $P : \psi \to \psi$ (**A little mistake in the lecture notes**) from (row) vectors to (row) vectors.

### Thm.1 (Stochastic matrix vs. Stationary distribution; Existence)

Assume that the state space $S$ is finite if not more assumptions are required.

Every stochastic matrix $P$ has at least one stationary distribution.

**Proof. (See my note 4)**

### Thm.2 (Markov Chain Convergence Theorem)

If $P$ is both aperiodic and irreducible, then

1. $P$ has exactly one stationary distribution $\psi^*$;
2. For any initial distribution $\psi_0$, we have

$$\|\psi_0 P^t - \psi^*\| \to 0 \; as \; t \to \infty \tag{7}$$

## 3.1.4 Example

### Marginal Distribution

For $X_t$, since it is a random variable, it follows a distribution

| $X_t$ | **0** | **1** |
|---|---|---|
| $P(X_t)$ | $P(X_t = 0)$ | $P(X_t = 1)$ |

Denote this distribution as $\psi_t = (P(X_t = 0), P(X_t = 1))$.

By the Law of toal probability, we can decompose the probability that $X_{t+1} = y, y \in S$:

$$P\{X_{t+1} = y\} = \sum_{x \in S} P\{X_{t+1} = y | X_t = x\} \cdot P\{X_t = x\} \tag{12}$$

In this example, when $X_{t+1} = 0$ and $X_{t+1} = 1$, we have

$$\begin{aligned} P\{X_{t+1} = 0\} &= \sum_{x \in S} P\{X_{t+1} = 0 | X_t = x\} \cdot P\{X_t = x\} \\ &= P\{X_{t+1} = 0 | X_t = 0\} \cdot P\{X_t = 0\} + P\{X_{t+1} = 0 | X_t = 1\} \cdot P\{X_t = 1\} \\ &= P(0,0) \cdot P\{X_t = 0\} + P(1,0) \cdot P\{X_t = 1\} \end{aligned} \tag{13}$$

and

$$\begin{aligned} P\{X_{t+1} = 1\} &= \sum_{x \in S} P\{X_{t+1} = 1 | X_t = x\} \cdot P\{X_t = x\} \\ &= P\{X_{t+1} = 1 | X_t = 0\} \cdot P\{X_t = 0\} + P\{X_{t+1} = 1 | X_t = 1\} \cdot P\{X_t = 1\} \\ &= P(0,1) \cdot P\{X_t = 0\} + P(1,1) \cdot P\{X_t = 1\} \end{aligned} \tag{14}$$

By $\psi_t$ and the Markov matrix $P$, the two equations above can be written in matrix form

$$\psi_{t+1} = \psi_t \cdot P \tag{15}$$

By repeating this $m$ times, we can move forward $m$ steps from $t$ into future $m + t$,

By iterating on the above matrix form, we get

$$\psi_{t+m} = \psi_t \cdot P^m \tag{16}$$

- where $P^m$ is the $m$-th power of $P$.

For a special case, if $\psi_0$ is the initial distribution from which $X_0$ is drawn, then $\psi_0 P^m$ is the distribution of $X_m$.

Let's repeat it

$$X_0 \sim \psi_0 \implies X_m \sim \psi_0 P^m \tag{17}$$

and more generally,

$$X_t \sim \psi_t \implies X_{t+m} \sim \psi_t P^m \tag{18}$$

### Irreducibility & aperiodicity

#### Two states communicates

Consider the Markov matrix $P$ and state space $S$ here.

Two states $0$ and $1$ of the Markov matrix $P$ are said to communicated with each other if there exist positive integer $n$ such that

$$P^{(n)}(0,1) > 0 \; and \; P^{(n)}(1,0) > 0 \tag{19}$$

Since $P(0,1) = \alpha$ and $P(1,0) = \beta$, for example, if $\alpha \neq 0$ and $\beta \neq 0$, we can find positive integer $1$ such that

$$\begin{gathered} P^{(1)}(0,1) = P(0,1) = \alpha > 0 \\ and \\ P^{(1)}(1,0) = P(1,0) = \beta > 0 \end{gathered} \tag{20}$$

So by definition of two-state communication, we can say that two states $0$ and $1$ communicate.

#### Irreducibility

By the same example, the Markov matrix $P$ is irreducibility if there exist positive integer $n$ such that

$$\begin{gathered} P^{(n)}(0,1) > 0 \; and \; P^{(n)}(1,0) > 0 \\ and \\ P^{(n)}(0,0) > 0 \; and \; P^{(n)}(1,1) > 0 \end{gathered} \tag{21}$$

For example, if $0 < \alpha < 1$ and $0 < \beta < 1$, the we can find positive integer $1$ such that

$$\begin{gathered} P^{(n)}(0,1) = \alpha > 0 \; and \; P^{(n)}(1,0) = \beta > 0 \\ and \\ P^{(n)}(0,0) = 1 - \alpha > 0 \; and \; P^{(n)}(1,1) = 1 - \beta > 0 \end{gathered} \tag{22}$$

By definition of communication, all states of the Markov matrix $P$ communicate.

By definition of irreducibility, the Markov matrix $P$ is irreducible

### Period and Aperiodicity

For example, if $0 < \alpha < 1$ and $0 < \beta < 1$, we have $0 < 1 - \alpha < 1$ and $0 < 1 - \beta < 1$. So we have

$$P^{(1)}(0,0) > 0 \; and \; P^{(1)}(1,1) > 0 \tag{23}$$

By definition of period, we can say that the period of the state $0$ is $1$, and the period of the state $1$ is $1$.

Since the period of every state of $P$ is $1$, by definition of aperiodicity, the Markov matrix $P$ is aperiodic.

## 3.2 Potential Contexts on slides

## 3.3 Manuscripts

# 4 Intro to Entry/Exit Model

## 4.1 Preparation

## 4.2 Potential Contexts on slides

## 4.3 Manuscripts

# 5 Carvalho & Grassi's Double-Pareto Theory

## 5.1 Preparation

### Assumption 1

**Firm size** evolves as a **Markov Chain** on the state space $\Phi = \{\varphi^s\}_{s=1,\cdots,S}$ with transition matrix

$$P = \begin{pmatrix} a+b & c & 0 & \cdots & \cdots & 0 & 0 \\ a & b & c & \cdots & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a & b & c \\ 0 & 0 & 0 & \cdots & 0 & a & b+c \end{pmatrix} \tag{1}$$

- where
  - $0 < a < 1$;
  - $0 < b < 1$;
  - $0 < c < 1$;
  - $a + b + c = 1$.

### Corollary 1

Define $\hat{\mu}_t = \frac{\mu_t}{M}$ for any $t$.

With a **continuum of firms**, or equivalently, when **aggregate uncertainty is absent**, $\epsilon_{t+1} = 0$:

$$\hat{\mu}_{t+1} = (\tilde{P}_t)'(\hat{\mu}_t + G) \tag{2}$$

- where

- $\tilde{P}_t$: the transition matrix $P$,
  - where the first $\tilde{s}(\mu_t) - 1$ rows are replaced by 0s;
    - where $\tilde{s}(\mu_t)$ is the threshold of the entry and exit rule when the variance-covariance of the $\epsilon_{t+1}$ is 0 (**???**).

## Corollary 2

Assume **Assumption 1** holds.

If the **potential entrants' firm size distribution is Pareto (i.e.,** $G_s = K_e(\varphi^s)^{-\delta_e}$**)**, then as $S \to \infty$, the **stationary firm size distribution converges point-wise to**:

$$\hat{\mu}_s = K_1(\frac{\varphi^s}{\varphi^{\bar{s^*}}})^{-\delta} + K_2(\frac{\varphi^s}{\varphi^{\bar{s^*}}})^{-\delta_e} \ for \ s \geq \bar{s^*} \tag{24}$$

- where
  - $\bar{s^*}$: the steady-state entry/exit thesholds for $S \to \infty$;
  - $\delta = \frac{\log(a/c)}{\log \varphi}$ (**??? Same as above, but ???**);
  - $K_1, K_2$: constants, independent of $s$.

## Basic idea

$$\mathbb{Q}\{\varphi^*_{s+1} = \varphi^{s+1} | \varphi^*_s = \varphi^s\} = P(\varphi^{s+1}, \varphi^s)1\{\varphi^s \geq \varphi^{s^*}\} + G(\varphi^{s+1})1\{\varphi^s < \varphi^{s^*}\} \tag{25}$$

## 5.2 Potential Contexts on slides

## 5.3 Manuscripts

# 6 My Conjecture, Uniqueness Proof and Supporting graphs from Simulation

## 6.1 Preparation*

### 6.1.1 Formal Expression of the conjecture

Given **Assumption 1**, and let $s^*$ be the threshold of the entry and exit rule at the stationary. If the potential entrants' firm size distribution is any distribution $G = (G(\varphi^1), G(\varphi^2), \cdots, G(\varphi^S))$

- Where (This means G can be any discrete distribution with $S$ values)
  - $0 < G(\varphi^s) \leq 1$; (**Important**)
  - $\sum_{s=1}^{S} G(\varphi^s) = 1$.

Then

1. as $S \to \infty$, the stationary firm size productivity distribution will **uniquely** converges point-wise to a **Zipf's distribution**, $\mu^* = (\mu_1, \mu_2, \cdots, \mu_S)$, where

$$\mu_s^* = \mathbb{P}\{\varphi = \varphi^s\} = C \cdot (\varphi^s)^{-\delta} \tag{6}$$

- where
  - $C$: a constant;
  - $\delta > 0$: pinned down by $a, c, \varphi$

2. This firm size stationary distribution is given by

$$\mu^* = \mu^* Q \tag{7}$$

- Where
  - $Q$ is the evolution Markov matrix followed by all firms at the stationary state;
    - follow the law of motion between incumbents and entrants:$$

# 6.1.2 Main idea of the proof***

If we want to prove that $Q$ only has one stationary distribution, then we should first prove its existence and then its uniqueness.

## Existence

To prove the existence of stationary distribution, by **Theorem 1**, we should prove the matrix $Q$ is a Markov matrix. It is easier to do that by using the definition of $Q$ and Markov matrix.

## Uniqueness

To prove its uniqueness, by **Theorem 2**, I should prove two features of the markov matrix $Q$: $Q$ is irreducible and aperiodic.

### Irreducibility

To prove $Q$ is irreducible, by definition of irreducibility, we should prove that all states communicate with each other, that is, for all $x, y \in \Phi$, we can find a positive integer $n$ such that

$$P^{(n)}(x,y) > 0 \; and \; P^{(n)}(y,x) > 0 \tag{26}$$

### Aperidicity

To prove $Q$ is aperiodic, by definition of aperiodicity, we should prove that for all $x \in \Phi$, we have

$$P(x,x) > 0 \tag{27}$$

## A trick comes from $Q$'s definition

Since $Q^{(y-x)}(x,y)$ can be represented as

$$Q^{(y-x)}(x,y) = \begin{cases} \sum_{s=1}^{S} [\sum_{i=1}^{S} Q^{(y-x-2)}(x,i) Q(i,s)] Q(s,y) & if \; x < y \\ Q(x,y) & if \; x = y \end{cases} \tag{10}$$

We notice that there exists one term $Q(x, x+1)Q(x+1, x+2)\cdots Q(y-1, y)$ in the expansional expression of $Q^{(y-x)}(x, y)$ when $x < y$; and $Q^{(y-x)}(x, y) = Q(x, y)$ when $x = y$.

Similarly, there exists one term $Q(y, y-1)Q(y-1, y-2)\cdots Q(x+1, x)$ involved in the expansional expression of $Q^{(y-x)}(y, x)$ when $y > x$; and $Q^{(y-x)}(x, y) = Q(x, y)$ when $x = y$.

Let $n = \{(y-x)|(x, y) \in \{1, \cdots, S\} \times \{1, \cdots, S\}\ and\ x \leq y\} = S - 1$.

Since other possible terms in the above expansional expressions are all non-negative, if we can prove that

1. $Q(s, s+1) > 0, \forall s \in \{1, \cdots, S-1\}$;
2. $Q(s, s-1) > 0, \forall s \in \{2, \cdots, S\}$.
3. $Q(s, s) > 0, \forall s \in \{1, \cdots, S\}$.

Then we can prove that for all $(x, y) \in \{(x, y) \in \{1, \cdots, S\} \times \{1, \cdots, S\}|x \leq y\}$, and $n = S - 1$, we have

1. $Q^n(x, y) > 0$ and;
2. $Q^n(y, x) > 0$.

By the definition of irreducibility, we know that the Markov matrix $Q$ is irreducible.

## 6.1.3 Preparation for the simulation

### A Setting up

1. state space: $s \in \{1, \cdots, S\}$;
2. Assumption 1: $P = \begin{pmatrix} a+b & c & 0 & \cdots & \cdots & 0 & 0 \\ a & b & c & \cdots & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & a & b & c \\ 0 & 0 & 0 & \cdots & 0 & a & b+c \end{pmatrix}_{S \times S}$;
3. The threshold of the entry and exit rule when $\mathbb{V}ar(\epsilon_{t+1}) = 0$: $s^* \in \{1, \cdots, S\}$;
4. Firm size state space: $\varphi^s \in \{\varphi^1, \cdots, \varphi^S\}$;
5. Firm size threshold: $\varphi^{s^*} \in \{\varphi^1, \cdots, \varphi^S\}$;
6. Stationary firm size distribution $\mu^* = (\mu_1^*, \cdots, \mu_S^*)$;
7. Law of motion for stationary firm size distribution between incumbents and entrants: $\mathbb{Q}\{\varphi_{s+1}^* = \varphi^{s+1}|\varphi_s^* = \varphi^s\} = P(\varphi^{s+1}, \varphi^s)1\{\varphi^s \geq \varphi^{s^*}\} + G(\varphi^{s+1})1\{\varphi^s < \varphi^{s^*}\}$.

### B Simulation: Steps

1. Define a function to generate Markov Matrix $Q$ for firm size distribution $\mu$ by using the Law of Motion, which considers Incumbents and Entrants' evolutions;
   - Incumbents's evolution follows a Markov matrix P and converges to a Pareto distribution at steady state;
   - Entrants' firm-level productivity distribution follows another distribution.

2. Define a function to calculate stationary firm size distribution by the iteration method based on Stationary Distribution Theorem 2.
3. Assign certain values to size of the state space $S$, threshold $s^*$ and other parameters.
4. Calculate 9 potential distribution (including Pareto (Zipf) distribution) of entrants' firm size, $G$ ;
5. For each potential distributioin of entrants', calculate stationary firm size distribution of all firms, $\mu^*$ by **Step 1 & 2 & 3** above, and plot each of them;
6. Take the log terms of both the distribution's probability mass $\mu_s^*$ and its support $\varphi^s$, and plot each of these log terms again;
7. Plot the Right-Hand tail of the log-term firm size distribution for each of 9 cases above.

### C Simulation: Expectations about the firm size distribution

#### Expectaction 1: Its shape depends on the threshold $s^*$

By the Law of Motion, if $s^*$ is larger, then we can observe that $\mu_s^*$ looks more like entrants' firm size distribution $G$; if $s^*$ is smaller, then we can observe that $\mu_s^*$ looks more like incumbents' firm size distribution, followed by Markov matrix $P$;

This expectation will be tested by the 1st & 2nd-type plotting graphs for each simulation (generated by Step 5 & 6). Verify

#### Expectation 2: the Right-hand tail of its log form will look like a straight line with a negative slope, as firm size increases.

Recall Proposition 3, stationary firm size distribution will converges to a Pareto distribution

$$\mu_s^* = \mathbb{P}\{\varphi^* = \varphi^s\} = C \cdot (\varphi^s)^{-\delta} \tag{9}$$

Take the log terms for both sides of the above equation, we will get

$$\log \mu_s^* = \log \mathbb{P}\{\varphi^* = \varphi^s\} = \log C - \delta \cdot \log(\varphi^s) \tag{10}$$

Since $\delta > 0$, then $\log \mu_s^*$ will be linear in $\log(\varphi^s)$ with a negative slope $-\delta$ on the right-hand tail of the firm productivity (size) distribution.

This expectation will be tested by the 2nd & 3rd-type plotting graphs for each simulation (generated by Step 6 & 7).

On the graphs of the right-hand side of the log-form plotting, we can observe a downward sloping straight line.

## 6.1.4 Simulation

# 6.2 Potential Contexts on slides

# 6.3 Manuscripts

# 7 What's next

## 7.1 Preparation

## 7.2 Potential Contexts on slides

## 7.3 Manuscripts