# Cournot Competition as a Continuous State Space Markov Chain 

Alexander Halsey<br>The University of Texas at Austin

In partial fulfillment of the requirements for graduation with the Dean's Scholars Honors Degree in Mathematics

Under the supervision of Dr. Joseph Neeman and Dr. Maxwell Stinchcombe

May 21, 2021


#### Abstract

The Cournot competition is a game in which two firms vie to produce the optimal quantity of a good. Perfectly rational and fully informed firms would produce the quantities given by the Nash equilibrium, the point at which neither firm could improve their payoff by changing their action. Although the Nash equilibrium for the Cournot competition is well understood, there are several proposed models describing how firms that are not fully informed or perfectly rational might still learn the Nash equilibrium. Two commonly used models are fictitious play and the successive best response strategy. I build on these by using a Markov chain, a model for the evolution of random systems, to capture the probabilistic behavior of imperfect firms. Most of the theory and applications of Markov chains deals with finite or countable state spaces. In order to make sense in the context of game theory, the theory of Markov chains on arbitrary state spaces must first be presented. I will provide the relevant results for general state space theory, then describe its novel applications for learning Nash equilibria in the Cournot model.


Keywords: Markov chain, general state space, Nash equilibrium, Cournot competition

## Contents

1 Background ..... 4
1.1 Game Theory ..... 4
1.2 Cournot Competition ..... 5
1.3 Learning Models in Repeated Games ..... 6
1.4 Markov Chains on Discrete State Spaces ..... 7
1.5 Markov Chains on General State Spaces ..... 8
2 Proposed Learning Models ..... 10
2.1 Imperfect Responses ..... 10
2.1.1 Defining the Markov Chain ..... 10
2.1.2 Irreducibility ..... 12
2.1.3 Recurrence and Aperiodicity ..... 14
2.1.4 Invariance ..... 15
2.1.5 Normally Distributed Errors ..... 18
2.1.6 Bounding the Stationary Measure ..... 21
2.1.7 Uniformly Distributed Errors ..... 27
2.2 Stochastic Fictitious Play ..... 29
2.2.1 Defining the Markov Chain ..... 29
2.2.2 Properties of the Chain ..... 30
2.2.3 Simulations ..... 31
3 Discussion ..... 34
4 Acknowledgements ..... 36
References ..... 37

## 1 Background

### 1.1 Game Theory

Game theory is the study of strategic interactions between rational decision-makers. A rational decision-maker always decides to maximize their expected utility, given their information. Each rational agent is aware that all the other agents they compete against are expected utility maximizers.

Definition 1.1. A game $\Gamma$ is a collection $\left(S_{i}, u_{i}\right)_{i \in I}$ where

- I is a non-empty set of agents
- for each $i \in I, S_{i}$ is $i$ 's set of possible strategies
- for each $i \in I, u_{i}: S \rightarrow \mathbb{R}$ is $i$ 's utility function, where $S:=\times_{i \in I} S_{i}$ is the product of the actions of all the players

The simplest type of game is a one-shot, simultaneous move game, in which each player $i$ simultaneously chooses an action $s_{i} \in S_{i}$ and receives a payoff $u_{i}(s), s:=\times_{i \in I} s_{i}$, that depends on the strategies of all the players in the game. I will use $s_{-i}$ to denote the strategy profile of all players except player $i$. Simple games can be represented using a payoff matrix, which specifies the payoffs for each player for every possible strategy profile. An example is the Prisoner's Dilemma, in which two players both have the choice to either cooperate or defect. Their payoffs can be represented by the following matrix:

$$
\begin{array}{ccc} 
& C & D \\
C & 2,2 & 0,4 \\
D & 4,0 & 1,1
\end{array}
$$

Here the first entry in each cell corresponds to the row player's utility, and the second to the column player's. If the players both play C , they each receive a payoff of 2 . If the row player plays $D$ and the column player plays $C$, then the row player receives a payoff of 4 and the column player receives a payoff of 0 .

Definition 1.2. A pure strategy Nash equilibrium for $\Gamma$ is a vector of strategies, $s^{*} \in S$ such that for all $s_{i} \in S_{i}, u_{i}\left(s_{i}^{*}, s_{-i}^{*}\right) \geq u_{i}\left(s_{i}, s_{-i}^{*}\right)$.

This definition captures the notion of mutual best responses: given the strategies of all the other agents, no agent can benefit by changing strategies. For example, in the Prisoner's Dilemma discussed above, $(D, D)$ is the unique Nash equilibrium. No player can benefit by switching to $C$ if their opponent is playing $D$, and any player will benefit by switching from $C$ to $D$ no matter what their opponent does. In this case, we say the strategy $D$ strictly dominates $C$. This is the source of the dilemma - $(C, C)$ is a better outcome for everyone compared to $(D, D)$, but no strictly rational actor would play $C$. Mixed strategy Nash equilibria also exist but will not be discussed in this paper.

### 1.2 Cournot Competition

Definition 1.3. Cournot competition is a game theoretic model describing quantity competition between two firms, where each firm $i$ decides to produce an amount $x_{i}$ of $a$ homogeneous good. Market inverse demand is set by some function depending on the sum $x_{1}+x_{2}$.

In this case, letting $x:=x_{1}+x_{2}$, I will take the inverse demand function $p$ to be $p(x)=A-B x$ for $A, B>0$. I will also assume that the firms are producing quantities $x_{1}, x_{2}$ at a constant marginal cost $C$, where $C<A$, which means their profit functions $\pi$ are given by

$$
\pi_{i}\left(x_{i}, x_{j}\right)=x_{i}\left(A-B\left(x_{i}+x_{j}\right)\right)-C x_{i}
$$

In the Cournot competition, each firm wants $i$ to choose the right quantity $x_{i}^{*}$ to produce to maximize its profit function. In this case, we can easily differentiate the profit functions and solve for the maximum to see that

$$
\begin{gathered}
\left(x_{1}^{*}, x_{2}^{*}\right)=\left(\frac{A-C-x_{2}^{*}}{2 B}, \frac{A-C-x_{1}^{*}}{2 B}\right) \\
x_{1}^{*}=\frac{A-C}{2 B+1} \\
x_{2}^{*}=\frac{A-C}{2 B+1}
\end{gathered}
$$

This vector $\left(x_{1}^{*}, x_{2}^{*}\right)$ is the Nash equilibrium for this problem, as both firms are maximizing their own profits given the quantity produced by the other. The symmetry is due to the
assumption that both firms face the same demand function and marginal cost. The latter assumption can be relaxed without complicating the solution. This computation also gives the "best response" correspondence. If firm $j$ produces $x_{j}$, then firm $i$ 's best response is to produce

$$
\begin{equation*}
\frac{A-C-x_{j}}{2 B} \tag{1}
\end{equation*}
$$

### 1.3 Learning Models in Repeated Games

Although the Nash equilibrium is a well-studied concept, there are several competing models for how agents repeatedly playing the same game might learn what the equilibrium is. In the real world, decisions about how much of a good to produce are not one-shot, simultaneous move games. Rather, they can be more realistically modeled as a sequence of one-shot, simultaneous move games, in which firms decide how much to produce at each time-step given the past history of production. We should expect them to gradually learn what the profit-maximizing quantity to produce is.

One learning model for Cournot competition is the successive best response model. In this model, each firm $i$ chooses the optimal response to the quantity produced by firm $j$ in the previous time-step. As this process continues, the firms gradually learn the equilibrium quantities. The proposition below requires $B>1 / 2$, which will be assumed for the rest of this paper.

Proposition 1. If $B>1 / 2$, the successive best response model for the Cournot competition specified above converges to the unique Nash equilibrium.

Proof. We can view the Nash equilibrium as the fixed point for the best response correspondence given above. The mapping $x_{i}^{t} \mapsto x_{i}^{t+2}$ can be calculated to be $x_{i}^{t+2}=k+\frac{x_{i}^{t}}{(2 B)^{2}}$, where $k=\frac{(2 B-1)(A-C)}{(2 B)^{2}}$. If $B>1 / 2$, this is a contraction mapping, so it converges to a unique fixed point $x_{i}^{*}$ by Banach's fixed point theorem. That is, the sequences of quantities produced at times $\{0,2,4, \ldots\}$ and at times $\{1,3,5, \ldots\}$ both converge to the same fixed point (by uniqueness), so the sequence of quantities produced by firm $i$ converges to $x_{i}^{*}$. This fixed point must be the Nash equilibrium quantity for firm $i$. Since both firms use the same adjustment process, the vector of quantities produced converges to $\left(x_{1}^{*}, x_{2}^{*}\right)$.

Another common learning model is deterministic fictitious play. Suppose firm $i$ produces
quantities $\left(x_{i}^{1}, \ldots, x_{i}^{T}\right)$ at times 1 through $T$. For all times $t>T, x_{i}^{t}=B r_{i}\left(\frac{1}{t} \sum_{n=1}^{t} x_{j}^{n}\right)$, where the best response function $B r_{i}\left(x_{j}\right)=\frac{A-C-x_{j}}{2 B}$ was given earlier. This model corresponds to a pattern where firms compete until time $T$, at which point they start best responding to the average of each other's past behavior.

Proposition 2. If the best response function is a contraction mapping, deterministic fictitious play converges to the unique Nash equilibrium.

Proof. See Deschamps [2].
These models assume that firms are fully aware of each other's behavior and that they always respond perfectly to one another. These assumptions are unrealistic for a model that attempts to capture behavior in the real world, and I will revisit them later.

### 1.4 Markov Chains on Discrete State Spaces

A Markov chain is a model for the evolution of a random process. The model is composed of the set of states the system or process can be in at any time, and a set of probabilities governing the transitions between states over time.

Definition 1.4. A discrete time Markov chain is a sequence $\Phi=\left\{\Phi_{n}, n \in \mathbb{Z}^{+}\right\}$, where $\Phi_{i}$ are random variables on a state space $X$ with the property that given $\Phi_{n}, \Phi_{n+1}$ is independent of $\Phi_{0}, \ldots, \Phi_{n-1}$.

In the simplest case, the state space $X$ contains only a finite or countable (i.e. discrete) number of states. The next state of the chain depends only on the current state of the chain, not its previous history. The values, or sample paths, of the chain $\Phi$ are points in the product space $\Omega=\Pi_{i=1}^{\infty} X_{i}$, where each $X_{i}$ is an exact copy of $X$. If we let $\mu$ be the initial distribution of the Markov chain, and $P_{x_{0}}\left(\Phi_{1}=x_{1}\right)$ be the probability of transitioning from the state $x_{0}$ to the state $x_{1}$, then the 'forgetfulness' property of the chain can be formally stated as follows. For every $n$, and for every sequence of states $\left\{x_{0}, \ldots, x_{n}\right\}$,

$$
P_{\mu}\left(\Phi_{0}=x_{0}, \Phi_{1}=x_{1}, \ldots, \Phi_{n}=x_{n}\right)=\mu\left(x_{0}\right) P_{x_{0}}\left(\Phi_{1}=x_{1}\right) P_{x_{1}}\left(\Phi_{1}=x_{2}\right) \ldots P_{x_{n-1}}\left(\Phi_{1}=x_{n}\right)
$$

$\Phi$ is called a time-homogeneous chain if the transition probabilities $P_{x_{i}}\left(\Phi_{1}=x_{i+1}\right)$ are independent of the times $i, i+1$. I will only consider discrete time, time homogeneous

Markov chains. For a time-homogeneous Markov chain, $P(x, y):=P_{x}\left(\Phi_{1}=y\right)$. Note that the forgetfulness property of the Markov chain can also be stated as

$$
P_{\mu}\left(\Phi_{n+1}=x_{n+1} \mid \Phi_{n}=x_{n}, \ldots, \Phi_{1}=x_{1}\right)=P\left(x_{n}, x_{n+1}\right)
$$

Definition 1.5. The matrix $P=\{P(x, y), x, y \in X\}$ is known as the transition matrix for a Markov chain, and satisfies the following for all $x, y$ :

- $P(x, y) \geq 0$
- $\sum_{z \in X} P(x, z)=1$

Definition 1.6. Given a transition matrix $P$, the $\boldsymbol{n}$-step transition matrix $P^{n}$ gives the distribution of states in the chain after $n$ steps. For $A \subseteq X$,

$$
P^{n}(x, A):=\sum_{y \in A} P^{n}(x, y)
$$

We often want to know whether it is possible for the chain to go from one state to another in a finite number of steps. To be precise, we say that if there is some $n$ such that $P^{n}(x, y)>0$, then $x \rightarrow y$. If $y \rightarrow x$ as well, then $x \leftrightarrow y$, or $x$ and $y$ communicate.

Proposition 3. Communication is an equivalence relation, so the equivalence classes $C(x)=\{y: x \leftrightarrow y\}$ cover the state space $X$, with $x \in C(x)$

Definition 1.7. If $C(x)=X$ for some state $x$, then the Markov chain is irreducible.

Irreducibility means that it is possible to go from any state in the chain to any other.

### 1.5 Markov Chains on General State Spaces

It is possible to construct Markov chains on continuous state spaces, and indeed much of the theory for discrete space chains generalizes nicely. I will define a Markov chain on a continuous state space following the convention of Meyn and Tweedie [5], building up from the transition probabilities. We will require the state space $X$ to be equipped with a countably generated $\sigma$-field, usually the Borel $\sigma$-field $\mathcal{B}(X)$. Note the similarity of the definitions and theorems below to those for discrete state space chains.

Definition 1.8. If $P=\{P(x, A), x \in X, A \in \mathcal{B}(X)\}$ satisfies

- for each $A \in \mathcal{B}(X), P(\cdot, A)$ is a non-negative measurable function on $X$
- for each $x \in X, P(x, \cdot)$ is a probability measure on $X$
then $P$ is known as a transition probability kernel.
Theorem 1. For any initial measure $\mu$ on $\mathcal{B}(X)$ and a transition probability kernel $P$ we can define a Markov chain $\Phi=\left\{\Phi_{1}, \Phi_{2}, \ldots\right\}$ on $\Omega=\prod_{i=1}^{\infty} X_{i}$ that is measurable with respect to $\mathcal{F}=\vee_{i=0}^{\infty} \mathcal{B}\left(X_{i}\right)$, and a probability measure $P_{\mu}$ on $\mathcal{F}$ such that $P_{\mu}(B)$ is the probability of the event $\Phi \in B$ for $B \in \mathcal{F}$; and for measurable $A_{i} \subseteq X_{i}, i=0,1, \ldots, n$ and any $n$, $P_{\mu}\left(\Phi_{0} \in A_{0}, \Phi_{1} \in A_{1}, \ldots, \Phi_{n} \in A_{n}\right)=\int_{y_{0} \in A_{0}} \cdots \int_{y_{n-1} \in A_{n-1}} \mu\left(d y_{0}\right) P\left(y_{0}, d y_{1}\right) \cdots P\left(y_{n-1}, A_{n}\right)$ Proof. See Meyn and Tweedie [5].

Irreducibility for Markov chains on discrete/countable state spaces deals with the hitting times of points in the space. Although sets can be reached with positive probability by a general state space Markov chain, we cannot say the same for individual points in the space, so we need a slightly different notion of irreducibility. Fortunately, we can use the analagous notion of $\psi$-irreducibility for a general state space chain. First, we define the relevant terms. Let $\Phi$ be a Markov chain evolving on a general state space $X$.

Definition 1.9. For any set $A \in \mathcal{B}(X)$ the first return time of $A$, denoted by $\tau_{A}$, is

$$
\min \left\{n \geq 1: \Phi_{n} \in A\right\}
$$

Definition 1.10. For any state $x \in X$ and set $A \in \mathcal{B}(X)$, the return time probability, written as $L(x, A)$, is $\mathbb{P}_{x}\left(\tau_{A}<\infty\right)$.
$L(x, A)$ gives the probability that starting from the state $x$, the chain reaches the set $A$ in a finite number of steps. With these definitions in mind, we use the framework for irreducibility described by Meyn and Tweedie [5].

Definition 1.11. A Markov chain $\Phi$ is called $\phi$-irreducible if there exists a measure $\phi$ on $\mathcal{B}(X)$ such that whenever $\phi(A)>0$, we have $L(x, A)>0$ for every $x \in X$.

This definition captures the idea that $\Phi$ hits "large" sets with positive probability, regardless of where it starts in the space. $\phi$-irreducibility is a relatively weak property; it does not guarantee that sets of $\phi$-measure 0 are avoided with probability 1 . Thus, we seek a stronger extension of $\phi$, given by the following theorem.

Theorem 2. If a Markov chain $\Phi$ is $\phi$-irreducible, then there exists a probability measure $\psi$ on $\mathcal{B}(X)$ such that

1. $\Phi$ is $\psi$-irreducible
2. for any other measure $\phi^{\prime}$, $\Phi$ is $\phi^{\prime}$-irreducible iff $\psi \succ \phi^{\prime}$
3. if $\psi(A)=0$, then $\psi\{y: L(y, A)>0\}=0$

The important features of $\psi$-irreducibility are that $\psi$ can be taken as a probability measure, and that "negligible" sets as measured by $\psi$ are avoided with probability $1 . \psi$ is sometimes known as a maximal irreducibility measure.

## 2 Proposed Learning Models

### 2.1 Imperfect Responses

### 2.1.1 Defining the Markov Chain

Both the fictitious play model and the successive best response strategy assume that the firms respond perfectly to one another at each timestep. Suppose instead that the firms are playing a successive best response strategy, where they choose the best response with probability $1-\epsilon$, where $\epsilon>0$. With probability $\epsilon$, they choose the "wrong" response. More precisely, at each time $t$, with probability $\epsilon$ firm $i$ decides what quantity to produce by sampling from some distribution $\nu$, where $\nu$ is either fixed or dependent only on the quantities produced by both firms in the previous timestep. This model captures the idea of firms acting either irrationally or being subject to shocks outside of their control. They make the rational decision most of the time, but every once in a while they make a random choice. I will assume that firms do not produce negative quantities, and that they will not produce more than $R:=\frac{A-C}{B}$ as this would guarantee they lose money given the demand function specified above. Thus, $\nu$ is always supported on the compact set $[0, R]$. There are several reasonable choices one can make for $\nu$. It can be uniform on $[0, R]$, or it can be normally distributed (truncated to $[0, R]$ ) such that at time $t, \mathbb{E}\left[\nu_{i}\right]=B r_{i}\left(x_{j}^{t-1}\right)$. The specific choice of $\nu$ will matter when simulating behavior. The theoretical results described
below require that $\nu$ be a strictly positive measure, meaning that $\nu$ assigns non-zero measure to every non-empty open subset of $[0, R]$.

Definition 2.1. The imperfect response transition kernel $P$ is defined as follows. Let $\epsilon>0$, and $A=\left(A_{1} \times A_{2}\right) \in \mathcal{B}\left([0, R]^{2}\right)$. Let the state at time $t$ be $\alpha=\left(x_{1}, x_{2}\right)$, and let $\nu_{\alpha_{i}}$ be a positive probability measure on $[0, R]$ assigned to $i \in\{1,2\}$ depending only on the current state $\alpha$. Then,

$$
\begin{gather*}
P\left(\left(x_{1}, x_{2}\right), A\right)= \\
\begin{cases}\left(1-\epsilon+\epsilon \nu_{\alpha_{1}}\left(A_{1}\right)\right)\left(1-\epsilon+\epsilon \nu_{\alpha_{2}}\left(A_{2}\right)\right), & \text { if } B r_{1}\left(x_{2}\right) \in A_{1}, B r_{2}\left(x_{1}\right) \in A_{2} \\
\left(1-\epsilon+\epsilon \nu_{\alpha_{1}}\left(A_{1}\right)\right) \epsilon \nu_{\alpha_{2}}\left(A_{2}\right), & \text { if } B r_{1}\left(x_{2}\right) \in A_{1}, B r_{2}\left(x_{1}\right) \notin A_{2} \\
\epsilon \nu_{\alpha_{1}}\left(A_{1}\right)\left(1-\epsilon+\epsilon \nu_{\alpha_{2}}\left(A_{2}\right)\right), & \text { if } B r_{1}\left(x_{2}\right) \notin A_{1}, B r_{2}\left(x_{1}\right) \in A_{2} \\
\epsilon \nu_{\alpha_{1}}\left(A_{1}\right) \epsilon \nu_{\alpha_{2}}\left(A_{2}\right), & \text { if } B r_{1}\left(x_{2}\right) \notin A_{1}, B r_{2}\left(x_{1}\right) \notin A_{2}\end{cases} \tag{2}
\end{gather*}
$$

where $\left(x_{i}, x_{j}\right) \mapsto\left(\operatorname{Br}_{i}\left(x_{j}\right), B r_{j}\left(x_{i}\right)\right)$ is the best response correspondence for the Cournot competition described above.

This transition kernel makes precise the "imperfect" behavior described directly above. By construction, the transition kernel is Markovian, since $\nu_{\alpha_{i}}$ and the best response correspondence depend only on the state $\alpha$.

Theorem 3. For any initial distribution $\mu$, the imperfect response transition kernel $P$ defines a Markov chain on $[0, R]^{2}$.

Proof. First we must check that $P$ meets the criteria for definition 1.8. The first is immediate. To show the second, fix $\alpha=\left(x_{1}, x_{2}\right) \in[0, R]^{2}$. Then,

$$
\begin{gathered}
P(\alpha, \cdot)=(1-\epsilon)^{2}+(1-\epsilon) \epsilon \int d \nu_{\alpha_{1}}+ \\
(1-\epsilon) \epsilon \int d \nu_{\alpha_{2}}+\epsilon^{2} \iint d \nu_{\alpha_{1}} d \nu_{\alpha_{2}}=(1-\epsilon)^{2}+2(1-\epsilon)(\epsilon)+\epsilon^{2}=1
\end{gathered}
$$

since $\nu_{\alpha_{i}}$ is a probability measure. By Theorem 1, the transition kernel defines a Markov chain with the given probability law.

Having shown the lemma, we can view the probabilistic Cournot adjustment process between two firms as a Markov chain evolving on the state space $[0, R]^{2}$. The state at time $t$ is the vector $\left(x_{1}^{t}, x_{2}^{t}\right)$ of quantities produced by firms 1 and 2 . The imperfect response
transition kernel gives a distribution over possible quantities produced at time $t+1$. Although this is a Markov chain for any initial distribution $\mu$, we will take $\mu=\delta_{\alpha}$ for some state $\alpha$, since real firms produce a fixed quantity, not a distribution over quantities.

I showed above that under the assumption that $B>1 / 2$, the successive best response strategy converges to the Nash equilibrium. The process just defined cannot remain at a single point, because even if the firms arrive at the Nash equilibrium, with probability $\epsilon$ they will deviate, which means they will almost surely move away from the Nash equilibrium at some time after arriving to it. However, the theory of general state space Markov chains will allow us to reach equally desirable conclusions about the imperfect response process.

### 2.1.2 Irreducibility

To show the imperfect response chain defined above is $\psi$-irreducible, we use T-chains. Although this may seem like more advanced a technique than is necessary, using T-chains allows us to take advantage of the numerous results presented by Meyn and Tweedie [5].

Definition 2.2. Let $a=\{a(n)\}$ be distribution on $\mathbb{Z}_{+}$, and let $\Phi$ be a Markov chain with transition kernel P. The $\mathbf{K}_{\mathbf{a}}$ chain with sampling distribution $\boldsymbol{a}$ is the Markov chain $\Phi_{a}$ with the transition kernel $K_{a}$ given by

$$
K_{a}(x, A):=\sum_{n=0}^{\infty} P^{n}(x, A) a(n) \quad x \in X, A \in \mathcal{B}(X)
$$

The sampled chain can be interpreted as the chain $\Phi$, sampled at time points drawn from $a$.

Definition 2.3. If a is a sampling distribution, and there exists a substochastic transition kernel $T$ satisfying

$$
K_{a}(x, A) \geq T(x, A) \quad x \in X, A \in \mathcal{B}(X)
$$

where $T(\cdot, A)$ is a lower semi-continuous function for any $A \in \mathcal{B}(X)$, then $T$ is called a continuous component of $K_{a}$.

Definition 2.4. If $\Phi$ is a Markov chain for which there exists a sampling distribution a such that $K_{a}$ possesses a continuous component $T$, with $T(x, X)>0$ for all $x \in X$, then $\Phi$ is a T-chain.

Theorem 4. If $\Phi$ is a T-chain and $L(x, O)>0$ for all $x$ and open sets $O \in \mathcal{B}(X)$ then $\Phi$ is $\psi$-irreducible.

With these definitions in mind, we show that the imperfect response chain is $\psi$-irreducible, for a large class of $\nu_{\alpha}$.

Theorem 5. Let $\Phi$ be the imperfect response chain defined in 2.1. For a state $\alpha=\left(x_{1}, x_{2}\right)$ and a set $A=A_{1} \times A_{2}, A \in \mathcal{B}(X)$, define the transition kernel $T$ as follows:

$$
T\left(\left(x_{1}, x_{2}\right), A\right)=\epsilon^{2} \nu_{\alpha_{1}}\left(A_{1}\right) \nu_{\alpha_{2}}\left(A_{2}\right)
$$

Assume that all the $\nu_{\alpha}$ are positive measures chosen so that $T(\cdot, A)$ is lower semi-continuous for all $x \in X$. Then, $\Phi$ is a $\psi$-irreducible T-chain.

Proof. Let $a=\delta_{1}$ be the sampling distribution used to define the $K_{a}$ chain. This is a trivial choice - $K_{\delta_{1}}$ has the same transition kernel as $\Phi$. Next, note that for any state $\alpha$,

$$
T(\alpha, \cdot)=\epsilon^{2} \iint d \nu_{\alpha_{1}} d \nu_{\alpha_{2}}=\epsilon^{2}
$$

This computation along with the assumptions made in the statement of the lemma show that $T$ is a continuous component of $K_{\delta_{1}}$. Clearly, since $K_{\delta_{1}}(x, A)=P(x, A)$, where $P$ is the imperfect response kernel, we conclude that $K_{\delta_{1}}(x, A) \geq T(x, A)$ for all $x \in X, A \in \mathcal{B}(X)$. Thus, the imperfect response chain is a T-chain. Finally, let $O=O_{1} \times O_{2}$ be an open set in $[0, R]^{2}$. Let $\alpha$ be any state in the chain. By construction, $P(\alpha, O) \geq \epsilon \nu_{\alpha_{1}}\left(O_{1}\right) * \epsilon \nu_{\alpha_{2}}\left(O_{2}\right)>0$. Then, $L(\alpha, O) \geq \mathbb{P}_{\alpha}\left(\tau_{O}=1\right)=P(\alpha, O)>0$. Thus, for all open sets $O$ and states $\alpha$, $L(\alpha, O)>0$, so by Theorem 4, the imperfect response chain is $\psi$-irreducible.

Proposition 4. For the imperfect response chain, the irreducibility measure $\psi$ assigns non-zero measure to every Borel set of non-zero Lebesgue measure.

Proof. Suppose there exists some set $A=A_{1} \times A_{2} \in \mathcal{B}\left([0, R]^{2}\right)$ of non-zero Lebesgue measure such that $\psi(A)=0$. Note that $\psi\{y: L(y, A)>0\}=0$, by Theorem 2. However, for all states $\alpha \in[0, R]^{2}, P(\alpha, A) \geq \epsilon \nu_{\alpha_{1}}\left(A_{1}\right) * \epsilon \nu_{\alpha_{2}}\left(A_{2}\right)>0$. Then, $\{y: L(y, A)>0\}=$ $[0, R]^{2}$, which would mean $\psi$ is trivial, a contradiction.

### 2.1.3 Recurrence and Aperiodicity

$\psi$-irreducibility, while useful, is quite an abstract property in the context of a Cournot competition. As argued before, we cannot guarantee this chain will remain at the Nash equilibrium forever. However, $\psi$-irreducibility allows us to show that the chain will hit any neighborhood of the Nash equilibrium infinitely many times. This seems like a powerful result, but $\psi$-irreducibility implies the chain will hit every open set in the space infinitely many times. Instead, we will focus on hitting times of open neighborhoods containing the Nash equilibrium. Ideally, the chain will get close to the Nash equilibrium quickly and keep returning there often relative to other states.

Definition 2.5. For a Markov chain $\Phi$ and a set $A \in \mathcal{B}(X)$, the occupation time is

$$
\eta_{A}:=\sum_{n=1}^{\infty} \mathbb{1}_{\left\{\Phi_{n} \in A\right\}}
$$

In this instance, we care about whether or not the imperfect response chain hits a given set $A$ infinitely many times, or equivalently that $\eta_{A}=\infty$.

Definition 2.6. $A$ set $A$ is called Harris recurrent if $Q(x, A):=\mathbb{P}_{x}\left(\eta_{A}=\infty\right)=1$ for all $x \in A$. A chain $\Phi$ is Harris recurrent if it is $\psi$-irreducible and every set $A \in \mathcal{B}(X)$ of positive $\psi$ measure is Harris recurrent.

To show the imperfect response chain is Harris recurrent, we use the following result without proof.

Theorem 6. $A \psi$-irreducible T-chain is Harris recurrent if and only if $P_{x}\{\Phi \rightarrow \infty\}=0$ for every $x \in X$, where $\{\Phi \rightarrow \infty\}$ means the chain $\Phi$ visits every compact set only finitely often.

Since the imperfect response chain evolves only on a compact set, by the theorem above it is Harris recurrent. As I showed earlier, every set $A \in \mathcal{B}\left([0, R]^{2}\right)$ with positive Lebesgue measure has positive $\psi$ measure, so if the chain starts at $\alpha \in A$, it returns to $A$ infinitely many times with probability one. This, combined with $\psi$-irreducibility, should imply the chain visits every open neighborhood containing the Nash equilibrium infinitely many times almost surely, regardless of where it starts.

Theorem 7. If a Markov chain is Harris recurrent, $Q(x, B)=1$ for every $x \in X$ and $B \in \mathcal{B}(X)$ such that $\psi(B)>0$.

Every neighborhood of the Nash equilibrium has positive $\psi$-measure, so by the theorem above, the imperfect response chain visits every neighborhood of the equilibrium infinitely many times almost surely. Of course, the theorem also implies that the chain visits every open neighborhood infinitely many times, so we look for stronger results using invariance.

### 2.1.4 Invariance

As a Markov chain $\Phi=\left\{\Phi_{n}\right\}$ evolves in time, we may expect it to approach a stable configuration, where stability means that for large $n$, the distribution of $\Phi_{n}$ is constant independently of $n$.

Definition 2.7. A stationary measure $\pi$ for a Markov chain with transition kernel $P$ evolving on a state space $X$ is a $\sigma$-finite measure $\pi$ on $\mathcal{B}(X)$ with the property that

$$
\pi(A)=\int_{X} \pi(d x) P(x, A), \quad A \in \mathcal{B}(X)
$$

The interpretation is that for a Markov chain $\Phi$ with initial distribution $\pi$, each $\Phi_{n}$ will be distributed according to $\pi$. Because the imperfect response chain is $\psi$-irreducible and recurrent, the following result applies.

Theorem 8. Suppose $\Phi$ is recurrent and $\psi$-irreducible. Then, there exists a unique (up to constant multiples) invariant measure $\pi$ on $\mathcal{B}(X)$ such that $\pi$ is equivalent to $\psi$ and satisfies for any $A, B \in \mathcal{B}(X)$ where $\psi(A)>0$,

$$
\pi(B)=\int_{A} \pi(d y) \mathbb{E}_{y}\left[\sum_{k=1}^{\tau_{A}} \mathbb{1}_{\left\{\Phi_{k} \in B\right\}}\right]
$$

This means that the stationary measure of $B$ is proportional to the amount of time spent in $B$ between successive visits of $A$, assuming the chain starts in $A$. This theorem guarantees existence of $\pi$, but it does not guarantee finiteness. Fortunately, because $\Phi$ is a T-chain evolving on a compact set, we can use the criterion of smallness.

Definition 2.8. $A$ set $C \in \mathcal{B}(X)$ is small if there exists an $m>0$, a constant $\delta \geq 0$ and a probability measure $\mu_{m}$ such that for all $x \in C, B \in \mathcal{B}(X)$,

$$
\begin{equation*}
P^{m}(x, B) \geq \delta \mu_{m}(B) \tag{3}
\end{equation*}
$$

A central result of the general state space theory is that for a $\psi$-irreducible chain, every Borel set of positive $\psi$ measure contains a small set. Another important result is that for a T-chain, compact sets are small. Because the imperfect response chain evolves on a compact set, we can use the following result to conclude that the imperfect response chain has a finite stationary measure.

Theorem 9. The invariant measure $\pi$ is finite if there exists a small set $C$ such that

$$
\sup _{x \in C} \mathbb{E}_{x}\left[\tau_{C}\right]<\infty
$$

If we take $C=[0, R]^{2}$, we immediately conclude that the unique stationary measure $\pi$ is finite for the imperfect response chain. Since $\pi$ is equivalent to $\psi$, and $\psi$ is positive on all Borel sets on non-zero Lebesgue measure, $\pi$ assigns positive probability to all Borel sets of non-zero Lebesgue measure. In particular, any neighborhood of the Nash equilibrium has positive $\pi$ measure. Before studying the long term dynamics of the chain, we define one more relevant term. In studying the long term behavior of the chain, we want to avoid scenarios where it cycles through states in some predictable fashion.

Definition 2.9. For a $\psi$-irreducible Markov chain, the largest $d$ such that there exist disjoint sets $D_{1} \ldots D_{d} \in \mathcal{B}(X)$ such that

- for $x \in D_{i}, P\left(x, D_{i+1}\right)=1, i=0 \ldots d-1 \bmod d$
- the set $N=\left[\cup_{i=1}^{d} D_{i}\right]^{c}$ is $\psi$-null
is known as the period of the chain. If $d=1$, the chain is aperiodic.
Proposition 5. The imperfect response chain is aperiodic.
Proof. To see why, recall that all the $\nu_{\alpha}$ defining the transition kernel $P$ are positive. If there were some cyclic decomposition of $\Phi$ into more than one set, then there would be a state $\alpha$, where $\alpha \in D, D=D_{1} \times D_{2}, D \in \mathcal{B}\left([0, R]^{2}\right)$ such that $P(\alpha, D)=0$. But, $P(\alpha, D) \geq \epsilon^{2} \nu_{\alpha_{1}}\left(D_{1}\right) \nu_{\alpha_{2}}\left(D_{2}\right)>0$. Thus, the chain is aperiodic.

Having shown aperiodicity, Harris recurrence, and the existence of a finite stationary measure, we use the following result.

Theorem 10. Suppose $\Phi$ is an aperiodic, Harris recurrent chain, with invariant measure $\pi$. If $\pi$ is finite, then for every initial state $x \in X$,

$$
\sup _{A \in \mathcal{B}(X)}\left|P^{n}(x, A)-\pi(A)\right| \rightarrow 0
$$

For the imperfect response chain, Theorem 10 implies that no matter where we start in the state space, $\Phi_{n}$ for large $n$ is distributed according to $\pi$. Intuitively, the chain spends most of its time in sets of relatively large $\pi$-measure. If the chain spends most of its time in sets containing the Nash equilibrium, then in some sense the competing firms have still "learned" the optimal quantities to produce, despite their imperfect response patterns. $\pi$ itself is difficult to calculate, and it should be evident that it will depend on the specific choice of $\nu_{\alpha}$. In the next section, we provide simulations for different choices of $\nu_{\alpha}$.

We can also get strong results about the speed of convergence to the stationary distribution. First, we define the relevant metric on probability distributions.

Definition 2.10. The total variation distance between two probability distributions $\mu$ and $\nu$ on a space $X$ is

$$
\|\nu(\cdot)-\mu(\cdot)\|=\sup _{A \in \mathcal{B}(X)}|\nu(A)-\mu(A)|
$$

With this definition in mind, we can apply the following result.
Theorem 11. Consider a Markov chain with stationary distribution $\pi$. Suppose the minorization condition (3) holds for some $n_{0} \in \mathbb{N}, \delta>0$, and probability measure $\mu$, in the special case $C=X$, i.e. the entire state space is small. Then,

$$
\begin{equation*}
\left\|P^{n}(x, \cdot)-\pi(\cdot)\right\| \leq(1-\delta)^{\left\lfloor n / n_{0}\right\rfloor} \tag{4}
\end{equation*}
$$

where $\lfloor r\rfloor$ is the greatest integer less than or equal to $r$.
See Roberts [6] for a proof. Since the entire state space is small for the imperfect response chain, convergence (in the sense of total variation distance) to the stationary distribution $\pi$ is both uniform over all starting states in the chain and relatively fast for all starting states.

### 2.1.5 Normally Distributed Errors

Assume the $\nu_{\alpha}$ for each firm are normally distributed about the best response to their competitor's last action. If the state at time $t$ is $\alpha=\left(x_{1}, x_{2}\right)$, then $\nu_{\alpha_{1}} \sim N\left(B r_{1}\left(x_{2}\right), \sigma\right)$ and $\nu_{\alpha_{2}} \sim N\left(B r_{2}\left(x_{1}\right), \sigma\right)$. Under this model, even if firms make the "wrong" choice with probability $\epsilon$, that choice is unlikely to be far from the optimal one. There is a small issue with the supports, as Gaussian distributions are not supported on compact sets. We can correct this by conditioning on whether or not each draw from the $\nu_{\alpha}$ is in $[0, R]^{2}$. In principle, this conditioning slightly alters the mean. In practice, if we take $\sigma$ small enough so that for the Nash equilibrium $\left(x_{1}^{*}, x_{2}^{*}\right),\left|x_{i}^{*} \pm 3 \sigma\right| \in[0, R]$ for each $i$, then the adjustment to the mean is negligible.

We need to check two properties before applying the results from the previous section. First, $\nu_{\alpha_{i}}=N\left(B r_{i}\left(x_{j}\right), \sigma\right)$ is strictly positive for all states $\alpha=\left(x_{1}, x_{2}\right)$, so the irreducibility results apply. Second, to show the imperfect response chain is a T-chain, we need to confirm that for all sets $A=A_{1} \times A_{2}, T(\cdot, A)=\epsilon^{2} \nu_{1}\left(A_{1}\right) \nu_{2}\left(A_{2}\right)$ is a lower semi-continuous function (see Proposition 5, and note that $\nu_{1}$ and $\nu_{2}$ depend on the state $(\cdot)$ ). This is immediate, because as $x_{n} \rightarrow x$, the best response correspondence $B r_{i}\left(x_{n}\right) \rightarrow B r_{i}(x)$, which means that the distributions $\nu_{\alpha_{i, n}}$, normally distributed about the best responses, converge on the measurable sets $A_{1}, A_{2}$. Thus, for this choice of $\nu_{\alpha}$, the imperfect response chain is $\psi$-irreducible, Harris recurrent, and uniformly ergodic with a finite stationary measure.

We can approximate ${ }^{1}$ the stationary distribution by allowing the Markov chain to run until its distribution is very close to stationary, then sampling states of the chain. We provide several examples below. In all cases, the parameters of the Cournot model were chosen as $A=100, B=2, C=5$, which means the Nash equilibrium quantity $x^{*}$ is 19 .

The simulations show that the stationary measure is always concentrated around the Nash equilibrium quantity, indicating that the firms do "learn" the optimal quantity to produce, though they continue to make mistakes. The effects of $\epsilon$ and $\sigma$ are as expected as well: as both $\epsilon$ and $\sigma$ increase, the response chain spends less time around the Nash equilibrium.

[^0]

Figure 1: Simulated stationary distributions of imperfect response process with error rate $\epsilon=0.1$, and standard deviations $\sigma=5$ (left) and $\sigma=2.5$ (right) for the normal distributions $\nu_{\alpha}$. For $\sigma=5, \pi\left(B_{1}\left(x^{*}\right)\right)=0.88$, and for $\sigma=2.5, \pi\left(B_{1}\left(x^{*}\right)\right)=0.92$.


Figure 2: Simulated stationary distributions of imperfect response process with error rate $\epsilon=0.25$, and standard deviations $\sigma=5$ (left) and $\sigma=2.5$ (right) for the normal distributions $\nu_{\alpha}$. For $\sigma=5, \pi\left(B_{1}\left(x^{*}\right)\right)=0.71$, and for $\sigma=2.5, \pi\left(B_{1}\left(x^{*}\right)\right)=0.81$.


Figure 3: Simulated stationary distributions of imperfect response process with error rate $\epsilon=0.50$, and standard deviations $\sigma=5$ (left) and $\sigma=2.5$ (right) for the normal distributions $\nu_{\alpha}$. For $\sigma=5, \pi\left(B_{1}\left(x^{*}\right)\right)=0.47$, and for $\sigma=2.5, \pi\left(B_{1}\left(x^{*}\right)\right)=0.63$.


Figure 4: Simulated stationary distribution of imperfect response process with error rate $\epsilon=$ 1 , and standard deviations $\sigma=5$ (top) and $\sigma=2.5$ (bottom) for the normal distributions $\nu_{\alpha}$. This reflects a scenario when the firms never pick the successive best response, and instead always draw their response from the normal distribution around the theoretical best response. The orange line represents a Gaussian with mean $x^{*}$ and standard deviation $\sigma$. For $\sigma=5, \pi\left(B_{1}\left(x_{i}^{*}\right)\right)=0.15$, and for $\sigma=2.5, \pi\left(B_{1}\left(x_{i}^{*}\right)\right)=0.30$.

### 2.1.6 Bounding the Stationary Measure

We can prove the stationary measure is relatively concentrated around the equilibrium for normally distributed $\nu_{\alpha}$ using the Wasserstein metric.

Definition 2.11. The $\boldsymbol{p}$-Wasserstein distance between two probability measures $\mu$ and $\nu$ on a metric space $(M, d)$ is

$$
W_{p}(\mu, \nu)=\left(\inf _{\gamma \in \Gamma(\mu, \nu)} \int_{M \times M} d(x, y)^{p} d \gamma(x, y)\right)^{1 / p}
$$

$\Gamma(\mu, \nu)$ is the set of couplings of $\mu$ and $\nu$, i.e. the set of measures on $M \times M$ whose marginal distributions are $\mu$ and $\nu$.

This metric is sometimes known as the "earth mover's distance", because if $\mu$ and $\nu$ are viewed as "piles" of mass, $W_{1}(\mu, \nu)$ represents the minimum cost required to turn $\mu$ into $\nu$. The distance term is used because the cost of moving from point $a$ to point $b$ is $d(a, b)$. The following well-known properties of the Wasserstein metric will be useful.

Remark 1. 1. For point masses $\delta_{x}, \delta_{y}, W_{1}\left(\delta_{x}, \delta_{y}\right)=|x-y|$.
2. For normal distributions $\nu_{1}=N\left(m_{1}, \sigma_{1}\right)$ and $\nu_{2}=N\left(m_{2}, \sigma_{2}\right)$,

- $W_{1}\left(\nu_{1}, \nu_{2}\right) \geq\left|m_{1}-m_{2}\right|$
- $W_{2}\left(\nu_{1}, \nu_{2}\right)^{2}=\left|m_{1}-m_{2}\right|^{2}+\sigma_{1}^{2}+\sigma_{2}^{2}-2 \sigma_{1} \sigma_{2}$

3. For any distributions $\mu, \nu, W_{1}(\mu, \nu) \leq W_{2}(\mu, \nu)$.
4. $W_{p}(\mu, \nu)=\left(\inf \mathbb{E}\left[d(X, Y)^{p}\right]\right)^{1 / p}$ where the infimum is taken over all joint distributions of the random variables $X, Y$ having marginal distributions $\mu, \nu$.

From 2 and 3, we conclude that for normal distributions with the same variance ( $\nu_{1}=$ $N\left(m_{1}, \sigma\right)$ and $\left.\nu_{2}=N\left(m_{2}, \sigma\right)\right), W_{1}\left(\nu_{1}, \nu_{2}\right)=\left|m_{1}-m_{2}\right| . W_{1}$ is the only metric that will be used in the rest of this paper, so we will simply refer to it as $W$.

Lemma 1. Consider the imperfect transition kernel with normally distributed errors $P$. Let $P_{1}$ and $P_{2}$ be $P$ restricted to the first and second coordinates. Then, for any states $\left(x_{1}, x_{2}\right)$ and $\left(y_{1}, y_{2}\right), W\left(\delta_{x_{1}} P_{1}, \delta_{y_{1}} P_{1}\right)+W\left(\delta_{x_{2}} P_{2}, \delta_{y_{2}} P_{2}\right) \leq \frac{1}{2 b}\left(W\left(\delta_{x_{1}}, \delta_{y_{1}}\right)+W\left(\delta_{x_{2}}, \delta_{y_{2}}\right)\right)$.

Proof. Note that $\delta_{x_{1}} P_{1} \sim(1-\epsilon) \delta_{B r_{1}\left(x_{2}\right)}+\epsilon N\left(B r_{1}\left(x_{2}\right), \sigma\right)$ and $\delta_{x_{2}} P_{2} \sim(1-\epsilon) \delta_{B r_{2}\left(x_{1}\right)}+$ $\epsilon N\left(B r_{2}\left(x_{1}\right), \sigma\right)$. Then,
$W\left(\delta_{x_{1}} P_{1}, \delta_{y_{1}} P_{1}\right)=W\left((1-\epsilon) \delta_{B r_{1}\left(x_{2}\right)}+\epsilon N\left(B r_{1}\left(x_{2}\right), \sigma\right),(1-\epsilon) \delta_{B r_{1}\left(y_{2}\right)}+\epsilon N\left(B r_{1}\left(y_{2}\right), \sigma\right) \leq\right.$
$(1-\epsilon) W\left(\delta_{B r_{1}\left(x_{2}\right)}, \delta_{B r_{1}\left(y_{2}\right)}\right)+\epsilon W\left(N\left(B r_{1}\left(x_{2}\right), \sigma\right), N\left(B r_{1}\left(y_{2}\right), \sigma\right)\right)=$ $(1-\epsilon)\left|B r_{1}\left(x_{2}\right)-B r_{1}\left(y_{2}\right)\right|+\epsilon\left|B r_{1}\left(x_{2}\right)-B r_{1}\left(y_{2}\right)\right|=$ $\left|B r_{1}\left(x_{2}\right)-B r_{1}\left(y_{2}\right)\right|=$ $\left|\frac{A-C-x_{2}}{2 b}-\frac{A-C-y_{2}}{2 b}\right|=$ $\frac{1}{2 b}\left|x_{2}-y_{2}\right|$
By the same logic, $W\left(\delta_{x_{2}} P_{2}, \delta_{y_{2}} P_{2}\right) \leq \frac{1}{2 b}\left|x_{1}-y_{1}\right|$. The conclusion follows by adding the two inequalities.

Lemma 2. For any distributions $\mu_{1}, \mu_{2}$ and $\nu_{1}, \nu_{2}$, $W\left(\mu_{1} P_{1}, \nu_{1} P_{1}\right)+W\left(\mu_{2} P_{2}, \nu_{2} P_{2}\right) \leq$ $\frac{1}{2 b}\left(W\left(\mu_{1}, \nu_{1}\right)+W\left(\mu_{2}, \nu_{2}\right)\right)$. That is, the result from the previous lemma holds for arbitrary distributions over states in the chain, not just point masses. $\mu_{i}$ and $\nu_{i}$ represent different distributions over firm i's choices.

Proof. The result from Lemma 1 generalizes to random measures $\delta_{X_{1}}, \delta_{X_{2}}, \delta_{Y_{1}}, \delta_{Y_{1}}$ where $X_{1}, X_{2}, Y_{1}, Y_{2}$ are random variables. To see why, simply note that $\delta_{X_{1}} P_{1} \sim(1-\epsilon) \delta_{B r_{1}\left(X_{2}\right)}+$ $\epsilon N\left(B r_{1}\left(X_{2}\right), \sigma\right)$, so the same calculations from the previous lemma are still valid. Thus,

$$
\begin{gathered}
W\left(\delta_{X_{1}} P_{1}, \delta_{Y_{1}} P_{1}\right)+W\left(\delta_{X_{2}} P_{2}, \delta_{Y_{2}} P_{2}\right) \leq \\
\frac{1}{2 b}\left(W\left(\delta_{X_{1}}, \delta_{Y_{1}}\right)+W\left(\delta_{X_{2}}, \delta_{Y_{2}}\right)\right)=\frac{1}{2 b}\left(\left|X_{1}-Y_{1}\right|+\left|X_{2}-Y_{2}\right|\right)
\end{gathered}
$$

Now, consider distributions $\mu_{1}, \mu_{2}$ and $\nu_{1}, \nu_{2}$. Recall that $W(\mu, \nu)=\inf \mathbb{E}|X-Y|$ where the infimum is taken over joint distributions of $X, Y$ having marginals $\mu$ and $\nu$. Since $[0, R]$ with the standard metric is a Polish space, there exist optimal couplings $\left(X_{1}, Y_{1}\right)$ and $\left(X_{2}, Y_{2}\right)$ such that $\mathbb{E}\left|X_{1}-Y_{1}\right|=W\left(\mu_{1}, \nu_{1}\right)$ and $\mathbb{E}\left|X_{2}-Y_{2}\right|=W\left(\mu_{2}, \nu_{2}\right)$ [7]. By what was shown above,

$$
\begin{gathered}
W\left(\delta_{X_{1}} P_{1}, \delta_{Y_{1}} P_{1}\right)+W\left(\delta_{X_{2}} P_{2}, \delta_{Y_{2}} P_{2}\right) \leq \frac{1}{2 b} W\left(\delta_{X_{1}}, \delta_{Y_{1}}\right)+W\left(\delta_{X_{2}}, \delta_{Y_{2}}\right)= \\
\frac{1}{2 b}\left(\left|X_{1}-Y_{1}\right|+\left|X_{2}-Y_{2}\right|\right)
\end{gathered}
$$

We take the expectation of both sides:

$$
\begin{aligned}
& \mathbb{E}\left[W\left(\delta_{X_{1}} P_{1}, \delta_{Y_{1}} P_{1}\right)+W\left(\delta_{X_{2}} P_{2}, \delta_{Y_{2}} P_{2}\right)\right] \leq \frac{1}{2 b} \mathbb{E}\left[\left|X_{1}-Y_{1}\right|+\left|X_{2}-Y_{2}\right|\right] \\
& \mathbb{E}\left[W\left(\delta_{X_{1}} P_{1}, \delta_{Y_{1}} P_{1}\right)+W\left(\delta_{X_{2}} P_{2}, \delta_{Y_{2}} P_{2}\right)\right] \leq \frac{1}{2 b}\left(W\left(\mu_{1}, \nu_{1}\right)+W\left(\mu_{2}, \nu_{2}\right)\right)
\end{aligned}
$$

The 1 -Wasserstein metric is jointly convex, so $\mathbb{E}[W(\cdot, \cdot)] \geq W(\mathbb{E}[\cdot], \mathbb{E}[\cdot])$. Thus,

$$
\begin{gathered}
W\left(\mathbb{E}\left[\delta_{X_{1}} P_{1}\right], \mathbb{E}\left[\delta_{Y_{1}} P_{1}\right]\right)+W\left(\mathbb{E}\left[\delta_{X_{2}} P_{2}\right], \mathbb{E}\left[\delta_{Y_{2}} P_{2}\right]\right) \leq \\
\mathbb{E}\left[W\left(\delta_{X_{1}} P_{1}, \delta_{Y_{1}} P_{1}\right)+W\left(\delta_{X_{2}} P_{2}, \delta_{Y_{2}} P_{2}\right)\right]
\end{gathered}
$$

The left hand side is equal to $W\left(\mu_{1} P_{1}, \nu_{1} P_{1}\right)+W\left(\mu_{2} P_{2}, \nu_{2} P_{2}\right)$ by definition of the expectation of a random measure, so combining the inequalities we conclude

$$
W\left(\mu_{1} P_{1}, \nu_{1} P_{1}\right)+W\left(\mu_{2} P_{2}, \nu_{2} P_{2}\right) \leq \frac{1}{2 b}\left(W\left(\mu_{1}, \nu_{1}\right)+W\left(\mu_{2}, \nu_{2}\right)\right)
$$

Lemma 3. For all $k \in \mathbb{N}$, and all distributions $\mu_{1}, \mu_{2}, \nu_{1}, \nu_{2}$,

$$
W\left(\mu_{1} P_{1}^{k}, \nu_{1} P_{1}^{k}\right)+W\left(\mu_{2} P_{2}^{k}, \nu_{2} P_{2}^{k}\right) \leq \frac{2 R}{(2 b)^{k}}
$$

Proof. The result from the previous lemma implies

$$
\begin{aligned}
& W\left(\mu_{1} P_{1}^{k}, \nu_{1} P_{1}^{k}\right)+W\left(\mu_{2} P_{2}^{k}, \nu_{2} P_{2}^{k}\right) \leq \frac{1}{2 b}\left(W\left(\mu_{1} P_{1}^{k-1}, \nu_{1} P_{1}^{k-1}\right)+W\left(\mu_{2} P_{2}^{k-1}, \nu_{2} P_{2}^{k-1}\right)\right) \leq \\
& \frac{1}{(2 b)^{2}}\left(W\left(\mu_{1} P_{1}^{k-2}, \nu_{1} P_{1}^{k-2}\right)+W\left(\mu_{2} P_{2}^{k-2}, \nu_{2} P_{2}^{k-2}\right)\right) \leq \ldots \leq \frac{1}{(2 b)^{k}}\left(W\left(\mu_{1}, \nu_{1}\right)+W\left(\mu_{2}, \nu_{2}\right)\right)
\end{aligned}
$$

For any distributions $\mu$ and $\nu$ on $[0, R]$,

$$
W(\mu, \nu)=\inf _{\gamma \in \Gamma(\mu, \nu)} \int_{[0, R] \times[0, R]} d(x, y) d \gamma(x, y) \leq \inf _{\gamma \in \Gamma(\mu, \nu)} \int_{[0, R] \times[0, R]} R d \gamma(x, y)=R
$$

Thus,

$$
\frac{1}{(2 b)^{k}}\left(W\left(\mu_{1}, \nu_{1}\right)+W\left(\mu_{2}, \nu_{2}\right)\right) \leq \frac{2 R}{(2 b)^{k}}
$$

We can exploit the symmetry of the problem in the two coordinates to simplify this inequality. Under the assumption that the best response functions (and therefore the equilibrium quantities) are the same for both firms, $P_{1}$ and $P_{2}$ and their stationary measures
$\pi_{1}$ and $\pi_{2}$ are the same. Take $\mu_{1}=\delta_{x^{*}}, \mu_{2}=\delta x^{*}, \nu_{1}=\pi_{1}$, and $\nu_{2}=\pi_{2}$. This allows us to compare the imperfect response chain starting from the equilibrium quantities $\left(x^{*}, x^{*}\right)$ to one starting from the stationary measures $\left(\pi_{1}, \pi_{2}\right)$. Then,

$$
W\left(\delta_{x^{*}} P_{1}^{k}, \pi_{1} P_{1}^{k}\right)+W\left(\delta_{x^{*}} P_{2}^{k}, \pi_{2} P_{2}^{k}\right) \leq \frac{2 R}{(2 b)^{k}}
$$

The two terms on the right hand side are the same, and $\pi_{i} P^{k}=\pi_{i}$, so

$$
W\left(\delta_{x^{*}} P_{i}^{k}, \pi_{i}\right) \leq \frac{R}{(2 b)^{k}}
$$

We can use the inequality above to provide a lower bound on the mass of the stationary distribution close to the equilibrium.

Theorem 12. If $B_{r}\left(x^{*}\right)$ is a ball of radius $r$ about the equilibrium, then for all $k \in \mathbb{N}$,

$$
\pi_{i}\left(B_{r}\left(x^{*}\right)\right) \geq(1-\epsilon)^{2 k}-\frac{R}{r} \frac{1}{(2 b)^{k}}
$$

Proof. Suppose both firms start at the Nash equilibrium. At each timestep, the probability that both firms remain at the equilibrium is $(1-\epsilon)^{2}$, since the Nash equilibrium is a fixed point of the best response function. Thus, $P^{k}\left(\left(x^{*}, x^{*}\right),\left(x^{*}, x^{*}\right)\right) \geq(1-\epsilon)^{2 k}$, or when viewed in one coordinate (still assuming the other firm initially produces the equilibrium quantity), $P_{i}^{k}\left(x^{*}, x^{*}\right) \geq(1-\epsilon)^{2 k}$. On the other hand, $W\left(\delta_{x^{*}} P_{i}^{k}, \pi_{i}\right) \leq \frac{R}{(2 b)^{k}}$.

Recall that we can think the 1-Wasserstein distance in terms of optimal transport, where $W\left(\delta_{x^{*}} P_{i}^{k}, \pi_{i}\right)$ is the minimal total cost of turning the first distribution into the second. Consider a ball of radius $r$ about $x^{*}$. For any coupling or transport plan $\gamma, \gamma(A \times B)$ is the amount of mass in $A$ that must get sent to $B$ (see Ambrosio and Gigli [1]). In order to turn $\delta_{x^{*}} P_{i}^{k}$ into $\pi_{i}$, we must transport a mass of at least $(1-\epsilon)^{2 k}-\pi_{i}\left(B_{r}\left(x^{*}\right)\right)$ from the Nash equilibrium to $\left(B_{r}\left(x^{*}\right)\right)^{c}$, which is a distance of at least $r$. To be precise, for any coupling $\gamma$ of $\delta_{x^{*}} P_{i}^{k}$ and $\pi_{i}$,

$$
\int_{[0, R] \times[0, R]}|x-y| d \gamma \geq \int_{x^{*} \times B_{r}\left(x^{*}\right)^{c}}|x-y| d \gamma \geq \int_{x^{*} \times B_{r}\left(x^{*}\right)^{c}} r d \gamma=r \cdot \gamma\left(x^{*} \times B_{r}\left(x^{*}\right)^{c}\right)
$$

Since the above is true for all couplings, it is true for the infimum (the Wasserstein distance). Then, for all $k$,

$$
\frac{R}{(2 b)^{k}} \geq W\left(\delta_{x^{*}} P_{i}^{k}, \pi_{i}\right) \geq r\left[(1-\epsilon)^{2 k}-\pi_{i}\left(B_{r}\left(x^{*}\right)\right)\right]
$$

Rearranging the inequality yields the desired result.

The bound provided is valid for all $k$, so to get the tightest lower bound we optimize over $k$, i.e. $\pi_{i}\left(B_{r}\left(x^{*}\right)\right) \geq \max _{k \in \mathbb{N}}(1-\epsilon)^{2 k}-\frac{R}{r} \frac{1}{(2 b)^{k}}$. We provide approximate lower bounds on $\pi_{i}\left(B_{r}\left(x^{*}\right)\right)$ for various choices of $\epsilon$ and $r$, given the values of $R$ and $b$ used for the simulations, and compare to the results of the simulations. Note that the lower bounds do not depend on the standard deviation $\sigma$ of the $\nu_{\alpha}$.

| $\epsilon$ | r | $\pi_{i}\left(B_{r}\left(x^{*}\right)\right)$ - lower bound | $\pi_{i}\left(B_{r}\left(x^{*}\right)\right)$ - simulation |
| :---: | :---: | :---: | :---: |
| 0.001 | 1 | 0.983 | 0.9985 |
| 0.001 | 2.5 | 0.985 | 0.9991 |
| 0.001 | 5 | 0.986 | 0.9997 |
| 0.01 | 1 | 0.875 | 0.989 |
| 0.01 | 2.5 | 0.886 | 0.994 |
| 0.01 | 5 | 0.895 | 0.997 |
| 0.05 | 1 | 0.552 | 0.939 |
| 0.05 | 2.5 | 0.589 | 0.969 |
| 0.05 | 5 | 0.626 | 0.985 |
| 0.10 | 1 | 0.302 | 0.877 |
| 0.10 | 2.5 | 0.356 | 0.932 |
| 0.10 | 5 | 0.393 | 0.967 |
| 0.15 | 1 | 0.150 | 0.822 |
| 0.15 | 2.5 | 0.198 | 0.902 |
| 0.15 | 5 | 0.235 | 0.952 |
| 0.20 | 1 | 0.061 | 0.763 |
| 0.20 | 2.5 | 0.094 | 0.869 |
| 0.20 | 5 | 0.131 | 0.936 |
|  |  |  |  |

Table 1: Behavior of the stationary distribution of the quantities produced by firm $i$ for varying error rates $\epsilon$. The errors are normally distributed with standard deviation $\sigma=5$. The third column represents a theoretical lower bound on the stationary mass of a ball of radius $r$ about the equilibrium $x^{*}$, while the fourth represents the simulated mass of that ball.

These results show that for small $\epsilon$, the stationary distribution for the imperfect response chain places a provably large amount of mass very close to the Nash equilibrium. These results are as expected; the limiting case $\epsilon=0$ corresponds to the successive best response dynamic, which converges to the Nash equilibrium. As $\epsilon$ grows, the theoretical bound drops off from the empirical results. However, in all cases the simulations show a large concentration around the equilibrium, indicating the firms learn to produce close to the optimal quantity in most time periods.

### 2.1.7 Uniformly Distributed Errors

We can also take the $\nu_{\alpha}$ to be uniform on $[0, R]$, which reflects a scenario where production shocks are completely random. Note the $\nu_{\alpha}$ are all the same, so for simplicity we will just write $\gamma$ for the uniform distribution on $[0, R]$.

We need to verify a few properties before using the results from previous sections. The uniform measure is strictly positive, and trivially $T\left(\cdot, A_{1} \times A_{2}\right)=\epsilon^{2} \gamma\left(A_{1}\right) \gamma\left(A_{2}\right)$ is a lower semicontinous function (since it is constant) for all sets $A=A_{1} \times A_{2}$. Thus, the imperfect response chain is $\psi$-irreducible, Harris recurrent, and uniformly ergodic with a finite stationary measure.

Similarly to the previous section, we can approximate the stationary measure by allowing the chain to run until its distribution is very close to stationary, then sampling states of the chain. The results are pictured below. Again, the parameters of the Cournot model were chosen as $A=100, B=2, C=5$, which means the Nash equilibrium quantity $x^{*}$ is 19 .


Figure 5: Simulated stationary distribution $\pi$ of imperfect response process with error rate $\epsilon=0.10$ for the uniform distribution $\gamma . \pi\left(B_{1}\left(x^{*}\right)\right)=0.80$.


Figure 6: Simulated stationary distribution $\pi$ of imperfect response process with error rate $\epsilon=0.25$ for the uniform distribution $\gamma . \pi\left(B_{1}\left(x^{*}\right)\right)=0.56$.


Figure 7: Simulated stationary distribution $\pi$ of imperfect response process with error rate $\epsilon=0.50$ for the uniform distribution $\gamma . \pi\left(B_{1}\left(x^{*}\right)\right)=0.27$.

The bias in the simulations occurs because the mean of the uniform distribution on $[0, R]$ is slightly greater than $x^{*}$. All the same theoretical bounds determined in the previous section still apply, since we can prove an analog of Lemma 1 for the uniform distribution.

Lemma 4. Consider the imperfect transition kernel with uniformly distributed errors $P$. Let $P_{1}$ and $P_{2}$ be $P$ restricted to the first and second coordinates. Then, for any states $\left(x_{1}, x_{2}\right)$ and $\left(y_{1}, y_{2}\right), W\left(\delta_{x_{1}} P_{1}, \delta_{y_{1}} P_{1}\right)+W\left(\delta_{x_{2}} P_{2}, \delta_{y_{2}} P_{2}\right) \leq \frac{1}{2 b}\left(W\left(\delta_{x_{1}}, \delta_{y_{1}}\right)+W\left(\delta_{x_{2}}, \delta_{y_{2}}\right)\right)$.

Proof. Note that $\delta_{x_{1}} P_{1} \sim(1-\epsilon) \delta_{B r_{1}\left(x_{2}\right)}+\epsilon \gamma$ and $\delta_{x_{2}} P_{2} \sim(1-\epsilon) \delta_{B r_{2}\left(x_{1}\right)}+\epsilon \gamma$. Then,

$$
W\left(\delta_{x_{1}} P_{1}, \delta_{y_{1}} P_{1}\right)=W\left((1-\epsilon) \delta_{B r_{1}\left(x_{2}\right)}+\epsilon \gamma,(1-\epsilon) \delta_{B r_{1}\left(y_{2}\right)}+\epsilon \gamma\right) \leq
$$

$$
\begin{gathered}
(1-\epsilon) W\left(\delta_{B r_{1}\left(x_{2}\right)}, \delta_{B r_{1}\left(y_{2}\right)}\right)+\epsilon W(\gamma, \gamma)= \\
(1-\epsilon)\left|B r_{1}\left(x_{2}\right)-B r_{1}\left(y_{2}\right)\right| \leq \\
\left|B r_{1}\left(x_{2}\right)-B r_{1}\left(y_{2}\right)\right|= \\
\left|\frac{A-C-x_{2}}{2 b}-\frac{A-C-y_{2}}{2 b}\right|= \\
\frac{1}{2 b}\left|x_{2}-y_{2}\right|
\end{gathered}
$$

By the same logic, $W\left(\delta_{x_{2}} P_{2}, \delta_{y_{2}} P_{2}\right) \leq \frac{1}{2 b}\left|x_{1}-y_{1}\right|$. The conclusion follows by adding the two inequalities.

Since this same result holds, the results from Lemmas 2 and 3 and Theorem 12 hold, so the theoretical lower bounds in table 1 still apply even if the errors are uniformly distributed.

### 2.2 Stochastic Fictitious Play

### 2.2.1 Defining the Markov Chain

We can also use a Markov chain approach to modify the fictitious play algorithm, which is less myopic and thus arguably more realistic than the successive best response dynamic. We expect firms to keep track of the past when competing, although they will still make mistakes. We can model this behavior by modifying the imperfect response Markov chain constructed in the previous section. We will prove a set of results for fictitious play similar to those proven for the best response dynamic.

Suppose two firms compete until some time $T$, then begin best responding to the average of each other's behavior over the last $T$ periods with probability $1-\epsilon$. With probability $\epsilon$, they choose a random response. To define this as a Markov chain requires some notational work, but the underlying idea and results are similar to the imperfect response process. The state space of the chain is the set of possible $T$-length production histories, i.e. the set of all vectors $\left(\left(x_{1}^{1}, x_{2}^{1}\right), \ldots,\left(x_{1}^{T}, x_{2}^{T}\right)\right)$ for $\left(x_{1}, x_{2}\right) \in[0, R]^{2}$. At each timestep, the "oldest" production quantities in the state drop off and are replaced by the new quantities produced at that timestep. This construction allows us to always keep track of the previous $T$ quantities produced.

Definition 2.12. The fictitious play with imperfect responses transition kernel $F$ is defined as follows. Let $\epsilon>0$ and $A=\left(A_{1}, A_{2}\right) \in \mathcal{B}\left([0, R]^{2}\right)$. Let the state at time $t$ be $\alpha=\left(\left(x_{1}^{t-T+1}, x_{2}^{t-T+1}\right), \ldots,\left(x_{1}^{t}, x_{2}^{t}\right)\right)$, and let $\nu_{\alpha_{i}}$ be a strictly positive probability measure on $[0, R]$ assigned to $i \in\{1,2\}$ depending only on the current state $\alpha$. Using the best response correspondence (1), let $\left(x_{i}^{t}\right)^{*}=B r_{i}\left(\frac{1}{T} \sum_{n=t-T+1}^{t} x_{j}^{n}\right)$. Then,

$$
\begin{gather*}
F\left(\alpha,\left(\left(x_{1}^{t-T+2}, x_{2}^{t-T+2}\right), \ldots,\left(x_{1}, x_{2}^{t}\right), A\right)=\right. \\
\begin{cases}\left(1-\epsilon+\epsilon \nu_{\alpha_{1}}\left(A_{1}\right)\right)\left(1-\epsilon+\epsilon \nu_{\alpha_{2}}\left(A_{2}\right)\right), & \text { if }\left(x_{1}^{t}\right)^{*} \in A_{1},\left(x_{2}^{t}\right)^{*} \in A_{2} \\
\left(1-\epsilon+\epsilon \nu_{\alpha_{1}}\left(A_{1}\right)\right) \epsilon \nu_{\alpha_{2}}\left(A_{2}\right), & \text { if }\left(x_{1}^{t}\right)^{*} \in A_{1},\left(x_{2}^{t}\right)^{*} \notin A_{2} \\
\epsilon \nu_{\alpha_{1}}\left(A_{1}\right)\left(1-\epsilon+\epsilon \nu_{\alpha_{2}}\left(A_{2}\right)\right), & \text { if }\left(x_{1}^{t}\right)^{*} \notin A_{1},\left(x_{2}^{t}\right)^{*} \in A_{2} \\
\epsilon \nu_{\alpha_{1}}\left(A_{1}\right) \epsilon \nu_{\alpha_{2}}\left(A_{2}\right), & \text { if }\left(x_{1}^{t}\right)^{*} \notin A_{1},\left(x_{2}^{t}\right)^{*} \notin A_{2}\end{cases} \tag{5}
\end{gather*}
$$

Note that even though the chain technically evolves on a $T \times 2$ ( $T$ pairs of quantities) dimensional state space, the first $T-1$ pairs at time $t+1$ are determined by simply shifting the state vector at time to the left by one, since it would not make sense for the history of the chain to change. Thus, $F\left(\alpha,\left(\left(x_{1}^{t-T+2}, x_{2}^{t-T+2}\right), \ldots,\left(x_{1}^{t}, x_{2}^{t}\right), A\right)=0\right.$ if $\left(x_{1}^{t-T+2}, x_{2}^{t-T+2}\right), \ldots,\left(x_{1}^{t}, x_{2}^{t}\right)$ are not the last T-1 entries of $\alpha$.

This chain is similar to the imperfect response chain - in fact, the imperfect response chain is the fictitious play chain for $T=1$. Most of the same results will apply: it is $\psi$-irreducible, a T-chain under reasonable assumptions, and uniformly ergodic with a finite stationary measure.

### 2.2.2 Properties of the Chain

Proposition 6. The fictitious play chain is a T-chain, as long as $\nu_{\alpha_{i}}(A)$ is a lower semicontinuous function of $\alpha_{i}$ for all $A \in \mathcal{B}([0, R])$.

Proof. Let $\alpha=\left(\left(x_{1}^{t-T+1}, x_{2}^{t-T+1}\right), \ldots,\left(x_{1}^{t}, x_{2}^{t}\right)\right)$ be the current state of the chain (i.e., the last $T$ production decisions made by each firm). Consider the sampling distribution $a(n)=\delta_{T}$, i.e. the $K_{a}$ chain has transition kernel $K_{a}(\alpha, A)=F^{T}(\alpha, A)$. Let $A=$ $\left(A_{1}^{1} \times A_{2}^{1}, \ldots, A_{1}^{T} \times A_{2}^{T}\right)$ be a set in $\mathcal{B}\left([0, R]^{2 \times T}\right)$. Define the substochastic transition kernel $T_{F}$ to be $T_{F}(\alpha, A)=\epsilon^{2 T} \nu_{\alpha_{1}^{1}}\left(A_{1}^{1}\right) \nu_{\alpha_{2}^{1}}\left(A_{2}^{1}\right) \cdots \nu_{\alpha_{1}^{T}}\left(A_{1}^{T}\right) \nu_{\alpha_{2}^{T}}\left(A_{2}^{T}\right)$. Here, $\nu_{\alpha_{i}^{t}}$ refers to the error
distribution $\nu_{\alpha}$ that the firm $i$ uses at time $t$. Finally, by construction $K_{a}(\alpha, A) \geq T(\alpha, A)$. $T_{F}$ is lower semi-continuous as long as $\left(\alpha_{i}\right)_{n} \rightarrow \alpha_{i}$ implies $\nu_{\left(\alpha_{i}\right)_{n}}(A) \rightarrow \nu_{\alpha_{i}}(A)$ for all $A \in \mathcal{B}([0, R])$, which was the same condition for the imperfect response chain.

Proposition 7. The fictitious play chain is a $\psi$-irreducible Harris chain.
Proof. Fix an open set $O=\left(O_{1}^{1} \times O_{2}^{1}, \ldots, O_{1}^{T} \times O_{2}^{T}\right)$. For any states $\alpha, L(\alpha, O) \geq$ $F^{T}(\alpha, O) \geq \epsilon^{2 T} \nu_{\alpha_{1}^{1}}\left(O_{1}^{1}\right) \nu_{\alpha_{2}^{1}}\left(O_{2}^{1}\right) \cdots \nu_{\alpha_{1}^{T}}\left(O_{1}^{T}\right) \nu_{\alpha_{2}^{T}}\left(O_{2}^{T}\right)>0$ since the $\nu_{\alpha}$ are strictly positive. By Theorem 4, this is enough to conclude the fictitious play chain is $\psi$-irreducible. The fictitious play chain is a $\psi$-irreducible T-chain evolving on a compact set, so by Theorem 6, it is Harris recurrent.

Because the fictitious play chain is Harris and $\psi$-irreducible, there exists an invariant measure $\pi$. By Theorem $9, \pi$ can be taken as a probability measure since the chain evolves on a compact set (meaning the entire state space is small). To conclude the chain converges to its stationary distribution, we need to show aperiodicity.

Proposition 8. The fictitious play chain is aperiodic.
Proof. We proceed by contradiction. Suppose we have some decomposition $D_{1}, \ldots D_{d}$ as in definition 2.9. Consider $D_{1}=\left(C_{1}^{1} \times C_{2}^{1}, \ldots, C_{1}^{T} \times C_{2}^{T}\right) . \quad F\left(D_{1}, D_{2}\right)=1$, so it must be true that $D_{2}=\left(C_{1}^{2} \times C_{2}^{2}, \ldots,[0, R]^{2}\right)$, since the $\nu_{\alpha}$ are assumed to be strictly positive measures. But then, since $D_{1} \subseteq D_{2}^{c}, D_{1} \subseteq\left(\left(C_{1}^{2}\right)^{c} \times\left(C_{2}^{2}\right)^{c}, \ldots, \emptyset\right)$. This cannot happen (by construction, no set in the state vector can ever be empty), so we conclude that there is no cyclic decomposition for $d>1$.

Thus, by Theorem 10, the fictitious play chain converges in the sense of total variation distance to its stationary distribution. By Theorem 11, this convergence is uniform across all starting states. By the same argument used in Proposition 4, $\psi$ places non-zero measure on each set of non-zero Lebesgue measure, which means $\pi$ does as well since $\pi$ and $\psi$ are equivalent (Theorem 8).

### 2.2.3 Simulations

We take the $\nu_{\alpha}$ normally distributed about the best responses, as was done in section 2.1.5. To simulate the stationary measure, we allow the chain to run for a long time and then
sample states. The parameters of the Cournot model were again chosen as $A=100, B=$ $2, C=5$, which means the equilibrium quantity is $x^{*}=19$. We provide simulations for a few different choices of the error rate $(\epsilon)$ and the amount of time periods considered $(T)$. In all cases, $\nu_{\alpha}$ are normally distributed with $\sigma=5$.



Figure 8: Simulated stationary distributions of fictitious play process with error rate $\epsilon=$ 0.10 for $T=5$ (left) and $T=20$ (right). In both cases, $\pi\left(B_{1}\left(x^{*}\right)\right)=0.915$.



Figure 9: Simulated stationary distributions of fictitious play process with error rate $\epsilon=$ 0.25 for $T=5$ (left) and $T=20$ (right). For $T=5, \pi\left(B_{1}\left(x^{*}\right)\right)=0.786$, and for $T=20$, $\pi\left(B_{1}\left(x^{*}\right)\right)=0.791$.


Figure 10: Simulated stationary distributions of fictitious play process with error rate $\epsilon=0.50$ for $T=5$ (left) and $T=20$ (right). For $T=5, \pi\left(B_{1}\left(x^{*}\right)\right)=0.571$, and for $T=20$, $\pi\left(B_{1}\left(x^{*}\right)\right)=0.578$.

We also provide a table with simulated values for $\pi\left(B_{r}\left(x^{*}\right)\right)$ for various choices of $\epsilon, r$, and $T$. These results show the stationary measure for the imperfect fictitious play chain is very concentrated about the equilibrium. As expected, the mass about the equilibrium decreases as the error rate $\epsilon$ increases. The results indicate a benefit to using fictitious play rather than successive best responses, although there is a sort of plateau effect where averaging over more periods does not necessarily improve the performance of the learning model. This requires further investigation - I believe it might be possible to show that the size of $\epsilon$ places an upper bound on how concentrated the stationary measure can be no matter how large the look-back window $T$ is. Nevertheless, this table shows that firms learn to play the Nash equilibrium very often under the imperfect fictitious play learning model.

| $\epsilon$ | T | $\pi_{i}\left(B_{1}\left(x^{*}\right)\right)$ | $\pi_{i}\left(B_{2.5}\left(x^{*}\right)\right)$ | $\pi_{i}\left(B_{5}\left(x^{*}\right)\right)$ |
| :---: | :---: | :---: | :---: | :---: |
| 0.001 | 1 | 0.9985 | 0.9991 | 0.9997 |
| 0.001 | 5 | 0.9992 | 0.9993 | 0.9997 |
| 0.001 | 20 | 0.9989 | 0.9992 | 0.9995 |
| 0.01 | 1 | 0.989 | 0.994 | 0.997 |
| 0.01 | 5 | 0.991 | 0.993 | 0.997 |
| 0.01 | 20 | 0.991 | 0.994 | 0.997 |
| 0.05 | 1 | 0.939 | 0.969 | 0.985 |
| 0.05 | 5 | 0.958 | 0.969 | 0.984 |
| 0.05 | 20 | 0.958 | 0.969 | 0.984 |
| 0.10 | 1 | 0.877 | 0.932 | 0.967 |
| 0.10 | 5 | 0.915 | 0.934 | 0.968 |
| 0.10 | 20 | 0.915 | 0.938 | 0.968 |
| 0.25 | 1 | 0.708 | 0.836 | 0.919 |
| 0.25 | 5 | 0.786 | 0.845 | 0.919 |
| 0.25 | 20 | 0.790 | 0.846 | 0.920 |
| 0.50 | 1 | 0.469 | 0.676 | 0.838 |
| 0.50 | 2.5 | 0.571 | 0.692 | 0.841 |
| 0.50 | 5 | 0.578 | 0.692 | 0.841 |

Table 2: Behavior of the stationary distribution of the quantities produced by firm $i$ under fictitious play for varying error rates $\epsilon$ and look-back windows $T$. The errors are normally distributed with standard deviation $\sigma=5$. Each column represents the simulated stationary mass of a ball of radius $r$ about the equilibrium $x^{*}$, for $r=1, r=2.5$, and $r=5$.

## 3 Discussion

We demonstrate that the learning process for a Cournot competition can be described as a Markov chain on a continuous state space. The different Markov chain constructions allowed us to study commonly used learning models while relaxing the assumption of perfect rationality. We show that even under these assumptions, firms engaged in a Cournot
competition will still 'learn' the Nash equilibrium, in a slightly weaker sense than what is usually intended. Although their imperfect responses prevent them from learning to always play the equilibrium, the simulations and ergodicity results show that the firms produce quantities very close to the equilibrium at most timesteps, and they learn to do so relatively quickly.

One interesting question is whether or not the Markov chain underlying a realistic learning model is time-stationary. If firms arrive close to the Nash equilibrium and begin to notice that producing the equilibrium quantity leads to higher profits, then they might discard whatever learning model they have been using in favor of always producing a fixed quantity close to or at the equilibrium. On one hand, the simulations show that this is basically what happens with learning models proposed in this paper. In addition, even if firms "know" the equilibrium, they are still prone to making mistakes or suffering production shocks. On the other hand, we could define this learning model differently as having different regimes, where as time goes on the probability of making the correct choice increases.

The Markov chain approach allows us to view this problem in even more generality. We can discard the assumption of fixed marginal cost and instead model the marginal cost each firm faces as a random variable. Since both firms should be vulnerable to the same cost shocks, we might model the marginal costs the two firms face as either the same random variable or a pair of correlated random variables. We could also assume the firms have incomplete memory in fictitious play, where they sample from some subset of their competitor's past actions rather than just seeing the past $T$ quantities. Kaniovsky and Young show strong convergence results for the fictitious play with sampling approach in $2 \times 2$ games [4], which may generalize to Cournot competition.

The game theory results in this paper were driven in large part by the work of Fudenburg and Levine [3], who give a thorough survey of fictitious play and present an outline for using Markov chains to study strategy profiles of iterated games. Several authors also show how Markov processes can be used to select between strict equilibria in iterated games. In future, the continuous state space theory presented here could be used to provide a different flavor of selection results for more general game theoretic models with more than one strict equilibrium.

## 4 Acknowledgements

I would like to thank Dr. Joe Neeman and Dr. Maxwell Stinchcombe for their guidance on this project and their patience in helping me. I am also grateful to my friends and family who have supported me through my time at UT. Finally, I would like to thank the Dean's Scholars Honors program for the opportunity to complete this project and for all the benefits, both personal and academic, it has provided me.

## References

[1] Luigi Ambrosio and Nicola Gigli. A User's Guide to Optimal Transport, pages 1-155. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.
[2] Robert Deschamps. An algorithm of game theory applied to the duopoly problem. European Economic Review, 6(2):187-194, 1975.
[3] Drew Fudenberg. The Theory of Learning in Games. MIT Press, Cambridge, MA, 1998.
[4] Yuri M. Kaniovski and H.Peyton Young. Learning dynamics in games with stochastic perturbations. Games and Economic Behavior, 11(2):330-363, 1995.
[5] Sean Meyn, Richard L. Tweedie, and Peter W. Glynn. Markov Chains and Stochastic Stability. Cambridge Mathematical Library. Cambridge University Press, 2 edition, 2009.
[6] Gareth O. Roberts and Jeffrey S. Rosenthal. General state space markov chains and mcmc algorithms. Probability Surveys, 1(none), Jan 2004.
[7] Feng-Yu Wang. Coupling and applications, 2010.


[^0]:    ${ }^{1}$ Code for all the simulations can be found at https://github.com/alexhalsey/cournot_simulations

