SUPPLEMENT TO "DEMAND FLUCTUATIONS IN THE READY-MIX CONCRETE INDUSTRY"

By Allan Collard-Wexler

APPENDIX A: Stochastic Algorithm

To compute the strategies associated with a Nash equilibrium of the dynamic game, I adapt the stochastic algorithm of Pakes and McGuire (2001) to the discrete action setup used in this paper since the state space has up to 1.4 million states.[1]

I define the *hit counter*, denoted $h(a_i, x)$, as the number of times the location $l = (a_i, x)$ has been visited by my algorithm. The hit counter is important since it allows me to keep track of the precision of the computation of $W(a_i, x)$ and $\Psi[a_i|x]$ using the discrete action stochastic algorithm (henceforth, DASA). Given a reward and transition cost function $r(\cdot)$ and $\tau(\cdot)$, as well as a demand transition matrix D, the DASA computes a solution to the dynamic game, characterized by the choice-specific value function $W(a_i, x)$, and the conditional choice probabilities Ψ.

ALGORITHM—Discrete Action Stochastic Algorithm (DASA): An iteration k of the DASA follows these steps:

1. Start in a location $l^k = \{a_i^k, x^k\}$, with values for W^k, Ψ^k, and h^k in memory.

2. Draw an action profile for other players $a_{-i}^k \sim \prod_{r \neq i} \Psi^k[a_r^k|x^k]$. Given the action profile $a^k = \{a_i^k, a_{-i}^k\}$, draw a state in the next period x^{k+1}:

$$(\text{S.1}) \qquad x^{k+1}|a^k \sim D[M^{k+1}|M^k] \prod_i \iota(x_i^{k+1}|a_i^k, x_i^k),$$

where $\iota(x_i^{k+1}|a_i^k, x_i^k)$ is the *updating* function, which updates the firm's state based on a firm's action and the firm's largest size in the past.[2]

3. Increment the hit counter (how often you have visited the state-action pair): $h^{k+1}(a_i^k, x^k) = h^k(a_i^k, x^k) + 1$.

---

[1]There are 10 firms, 7 possible states per firm, and, in the most complex model, 50 demand states. I reduce the size of the state from $10^7 \times 50$ to 1.4 million by using the assumption of exchangeability described by Gowrisankaran (1999).

[2]Later in the paper, I make the firm's previous state relevant to the transition cost. Specifically, if the firm's state is $x_i^k = \{x_i^{C,k}, x_i^{P,k}\}$, that is, the current size $x_i^{C,k}$ and the largest size in the past $x_i^{P,k}$, then the updating function $\iota(x_i^{k+1}|a_i^k, x_i^k)$ is

$$x_i^{k+1} = \begin{cases} \{a_i^k, a_i^k\}, & \text{if } a_i^k \geq x_i^{P,k}, \\ \{a_i^k, x_i^{P,k}\}, & \text{if } a_i^k < x_i^{P,k}. \end{cases}$$

4. Compute the value R of the action as

$$(S.2) \quad R = r(a_i^{k+1}, x^{k+1}) - \tau(a_i^{k+1}, x_i^k)$$

$$+ \beta \sum_{j \in A_i} W^k(j, x^{k+1}) \Psi^k[j|x^{k+1}] + \beta E(\varepsilon|x^{k+1}, \Psi^k),$$

where $E(\varepsilon|x^{k+1}, \Psi^k) = (\gamma - \sum_{j \in A} \ln(\Psi^k[j|x^{k+1}]) \Psi^k[j|x^{k+1}])$ (where γ is Euler's constant).

5. Update the W function:

$$(S.3) \quad W^{k+1}(a_i^k, x^k) = \alpha[a_i^k, x^k]R + (1 - \alpha[a_i^k, x^k])W^k(a_i^k, x^k),$$

where $\alpha = \frac{1}{h^{k+1}(a_i^k, x^k)}$.[3]

6. Update the policy function Ψ for state x^k:

$$(S.4) \quad \Psi^{k+1}[a_i^k|x^k] = \frac{\exp(W^{k+1}(a_i^k, x^k))}{\sum\limits_{j \in A_i} \exp(W^{k+1}(j, x^k))}$$

for all actions $a_i^k \in A$.

7. Draw a new action $a_i^{k+1} \sim \Psi^{k+1}[\cdot|x^{k+1}]$.

8. Check the stopping rule.[4] If it is not satisfied, update the current location to $l^{k+1} = \{a_i^{k+1}, x^{k+1}\}$, increment k to $k+1$, and return to step 1.

The stopping rule for this algorithm is based on Fershtman and Pakes (2012), which compares the W function to a simulated average based on rewards from steps 2 and 4 for states that are recurrent. If the W function is exact, then the squared difference between these two objects (weighted by the ergodic distribution) can be accounted for by simulation error. The stopping rule is presented in the next section.

---

[3]The main problem with the stochastic algorithm is (1) making sure the entire state space is searched, (2) ensuring fast learning about the W function at the start of the algorithm, and (3) making sure that the convergence properties of the algorithm are satisfied. First, I initialize the starting W using fairly high values so that the algorithm visits all states before lowering the estimate of W. Second, at the start of the algorithm, I use $\alpha = 1/\sqrt{h^{k+1}(a_i^k, x^k)}$ to ensure that initially inaccurate W's get updated quickly. As well, I reset the hit counter after 20 million iterations to ensure that the first rounds of updates are down-weighted. Third, in the final stage of the algorithm, I switch to the $\alpha = 1/h^{k+1}(a_i^k, x^k)$ update rule, which satisfies the convergence properties of stochastic approximation algorithms described in Powell (2007 p. 216). However, in the context of a game, this condition on α may not be enough to guarantee convergence.

[4]Since the stopping rule is computationally intensive relative to a single iteration of the DASA, it is better to check the stopping rule only every several million iterations.

The initial values of W in the stochastic algorithm are important, since if I initialize $W(a, x)$ with a high value, the algorithm might get trapped at this state. To find initial values of W, I use value iteration, in which I simulate the expectation via Monte Carlo.

A.1. *Discrete Action Stochastic Algorithm: Termination Criteria*

The stopping rule is based on the fact that if I have the "correct" W function, then it satisfies the Bellman equation. However, it is computationally expensive to calculate the W function exactly; instead, we can approximate the value function using forward simulation. Consider the locations $R \subset A_i \times X$ defined as the state-action pairs visited in the last 1 million iterations (keep a hit counter that tracks the last 1 million iterations denoted $rh(l)$).

ALGORITHM—Fershtman–Pakes Stopping Rule (FPStop): For all locations $l = \{a_i, x\} \in A_i \times X$ which have been visited in the last 1 million iterations:
1. Compute an approximation to the W function using a one step forward simulation (denoted $\tilde{W}$). For $q = 1, \ldots, Q$ (I use $Q = 10{,}000$):
  (a) Draw a state tomorrow $x^{q'}$ given location $l = \{a_i, x\}$.
  (b) Get rewards:

$$(\text{S.5}) \qquad R^q = r\big(a^q | x^{q'}, \theta\big) + \tau\big(a_i^q | x_i, \theta\big)$$
$$+ \beta \sum_{j \in A} W\big(j, x^{q'}\big) \Psi\big[j | x^{q'}\big]$$
$$+ \beta \bigg(\gamma - \sum_{j \in A} \ln\big(\Psi\big[j | x^{q'}\big]\big) \Psi\big[j | x^{q'}\big] \bigg).$$

  (c) Compute the approximation to the W function:

$$(\text{S.6}) \qquad \tilde{W}(l) = \frac{1}{Q} \sum_{q=1}^{Q} R^q.$$

2. Compute the difference in value functions weighted by the recent hit counter rh:

$$(\text{S.7}) \qquad \tau = \frac{1}{\displaystyle\sum_l rh(l)} \sum_l rh(l) * \big(\tilde{W}(l) - W(l)\big)^2.$$

If the test statistic τ is small enough, then we can argue that we have a good approximation. In practice, I use the recent hit counter weighted R^2 between $\tilde{W}(l)$ and $W(l)$ being greater than 0.999. This usually happens after as little as 50 million iterations, and it is usually more efficient to run the DASA for 150

million iterations (i.e., 15 minutes), which leads to a W function that satisfies the FPStop criteria. Furthermore, in this application, there are only about 3000 state-action pairs (where the action is not 0) that are visited in the last 1 million iterations. Thus, the ergodic class R is quite small compared to the size of the entire state space.

APPENDIX B: MODIFIED DASA TO COMPUTE THE GAMMA FUNCTION

I use a modified DASA to compute the Γ function. The two differences are that (i) I shut down the policy function update in the DASA, and (ii) I compute the net present value of the components of rewards rather than the rewards themselves (which would require me to have information on the parameters θ).[5]

ALGORITHM—Γ-Compute Discrete Action Stochastic Algorithm (GC-DASA): An iteration k of the GC-DASA is given by the following steps:

1. Start in a location $l^k = \{a_i^k, x^k\}$ with values for Γ^k and h^k in memory.

2. Draw an action profile $a^k | a_i^k \sim 1(a_i = a_i^k) \prod_{-i} \hat{P}[a_{-i}^k | x^k]$ and a state in the next period x^{k+1} given action profile a^k:

$$(S.9) \qquad x^{k+1} | a^k \sim \hat{D}[M^{k+1} | M^k] \prod_i \iota\left(x_i^{k+1} | a_i^k, x_i^k\right),$$

where $\iota(x_i^{k+1} | a_i^k, x_i^k)$ is the *updating* function, which updates the firm's state based on a firm's action and the firm's largest size in the past.

3. Increment the hit counter (how often you have visited the state-action pair): $h^{k+1}(a_i^k, x^k) = h^k(a_i^k, x^k) + 1$.

4. Compute the ith component of payoffs R^i of the action a_i^k as

$$(S.10) \qquad R^i = r^i\left(a_i^k, x^{k+1}\right)$$
$$+ \beta \sum_{j \in A} \Gamma^{k,i}\left(j, x^{k+1}\right) \hat{P}[j | x^{k+1}].$$

5. Update the Γ-function:

$$(S.11) \qquad \Gamma^{k+1,i}\left(a_i^k, x^k\right) = \alpha R^i + (1 - \alpha)\Gamma^{k,i}\left(a_i^k, x^k\right),$$

[5]I could have computed the Γ^{P^μ} using forward simulation, that is,

$$(S.8) \qquad \Gamma^{P^\mu}(a_i, s) \approx \frac{1}{K} \sum_{k=1}^{K} \sum_{t=0}^{\infty} \beta^t \vec{\rho}(a_{itk}, x_{tk}),$$

where the sequence of states x^{tk} can be simulated using demand transition process $\hat{D}$ and the choice probabilities for firms $\hat{P}$. However, there are about 350,000 states and 4 actions; thus, I would need to do this forward simulation 1.4 million times the number of simulation draws K.

where $\alpha = \frac{1}{h^{k+1}(a_i^k, x^k)}$.

6. Update current location to $l^{k+1} = \{a_i^k, x^{k+1}\}$.

The stopping rule is that of Fershtman and Pakes (2012).

APPENDIX C: MARKET FIXED EFFECTS

C.1. *Conditional Choice Probability Estimation*

In the main model, I use a market categories model which is meant to mimic the inclusion of market-fixed effects. These market-fixed effects are critical to the estimation of the model since persistent market-level differences in profitability lead to upward bias on the effect of competition. This bias, especially when it induces positive effects of competition, leads to very aberrant industry dynamics, such as having a market flip between 0 and 10 plants due to a positive externality due to competition. The goal of this section is to motivate the use of market-category effects based on the average number of firms in a market over time, and explain why other plausible corrections for market-fixed effects using average construction employment or the number of plants in a pre-period, do not give the right answer.

I consider the following different specifications of the market-category effects:

(a) *No Market Effects*.

(b) *Average Number of Firms in Market (rounded to nearest integer)*. In the main estimates of the model, I use the average number of firms in the market rounded to the nearest integer. However, this approach suffers from an endogeneity problem. To put it most clearly, consider the following dynamic, two firm model of the type

$$(S.12) \qquad a_{it} = \alpha a_{-it} + \beta a_{it-1} + \varepsilon_{it}.$$

If I include a_{it+1} in the above regression, then I am including an endogenous regressor since a_{it+1} is a function of a_{it} which, in turn, depends on ε_{it}, and more broadly, on the entire history of $\varepsilon_{i\tilde{t}}$ for $\tilde{t} < t$.

(c) *Average Number of Firms in Market in Years Before This One (rounded to nearest integer)*. However, if I include lagged a_{it-2}, then this is not an endogenous regressor, since there is no dependence on a_{it-2} except through a_{it-1}, which is already included in the regression. The only issue with using the average number of firms in the market in previous years is that a market can switch categories over time, which makes for a more difficult state space to deal with, which is the reason that I do not use these market-category controls in the main part of the paper.

(d) *Average Number of Firms in the 1976–1983 Period, With Data on the Post-1983 Period*. Notice that, for this model, I am using the early period to condition

the number of firms in the market. This is a version of model (c), but the pre-period on which I condition does not change within a market.

(e) *Average Construction Employment*. Here I classify markets by the average level of construction employment from 1976 to 1999. This is an exogenous classification scheme since it does not depend on what ready-mix concrete firms are doing.

(f) *Market-Fixed Effects (Conditional Logit)*.

Table S.I presents estimates from the binary logit model of entry and exit for specifications (a)–(f). I have chosen the binary logit model since it allows me to use the conditional logit with market-fixed effects.[6] Column (a) shows estimates without market-category controls (henceforth referred to as no market effects), column (f) shows estimates with market-fixed effects (henceforth referred to as market-fixed effects), while columns (b)–(e) show different market-category controls. Columns (b) and (c), that is, with market controls based on the average number of plants and the average number of plants in the periods before this one, are similar to the market-fixed effect estimates in column (f). Likewise, columns (d) and (e) show estimates that are more similar to the no market effects estimates in column (a).

The effects of past plant size on activity are fairly similar in all of these estimates, with smaller effects of plant size in the market-fixed effect specifications (f), (b), and (c) than in the no market effect specifications (a), (d), and (e). Unobserved heterogeneity between markets is loaded onto variable indicating state dependence, such as past plant size. The effect of log construction employment is higher at 0.133 to 0.099 in the no market effect models (a), (d), and (e) than in the market-fixed effect estimates, which have estimates from 0.033 to −0.034. These higher effects of demand are due to the fact that firms are far more likely to react to cross-sectional differences in demand (which are more likely to be persistent) than to year-to-year changes in demand. Likewise, the effect of the second competitor (which will be representative of the effect of competition more broadly) varies from −0.074 to 0.003 in the no market effect columns (a), (d), and (e), but ranges from −0.635 to −0.529 in the market-fixed effect columns (f), (b), and (c). This is indicative of the fact that unobserved differences in the profitability of a market will be correlated with the number of plants in the market.

There are two main conclusions from the table that are relevant for my choice of market categories. First, the market categories based on either the average number of firms (b) or the average number of firms in all periods before today (c) do a good job in mimicking true market-fixed effects. However,

---

[6]Technically, I can also use a multinomial conditional logit, but the number of categories I need to condition on becomes fairly large. As well, I am not presenting marginal effects here, since the conditional logit does not estimate the market-fixed effects.

TABLE S.I

MARKET EFFECTS IN THE BINOMIAL LOGIT REGRESSION OF ENTRY AND EXIT[a]

Dependent Variable: Activity	(a)	(b)	(c)	(d)	(e)	(f) Conditional Logit
Log County Construction Employment	0.133***	−0.034**	−0.034***	0.129***	0.099***	0.033
	(0.011)	(0.011)	(0.010)	(0.015)	(0.019)	(0.023)
First Competitor	−1.403***	−1.805***	−1.748***	−1.306***	−1.421***	−2.002***
	(0.052)	(0.051)	(0.048)	(0.066)	(0.052)	(0.043)
Second Competitor	0.003	−0.529***	−0.553***	−0.074	−0.008	−0.635***
	(0.036)	(0.037)	(0.036)	(0.047)	(0.037)	(0.030)
Third Competitor	0.026	−0.359***	−0.384***	−0.071	0.027	−0.394***
	(0.044)	(0.044)	(0.043)	(0.058)	(0.044)	(0.036)
Log Competitors Above 4	0.022	−0.118***	−0.170***	−0.001	0.035	−0.187***
	(0.029)	(0.028)	(0.028)	(0.040)	(0.029)	(0.025)
Small	5.889***	5.703***	5.720***	5.977***	5.887***	5.585***
	(0.037)	(0.035)	(0.035)	(0.047)	(0.037)	(0.025)
Small, Medium in Past	5.665***	5.388***	5.393***	5.707***	5.657***	5.220***
	(0.048)	(0.045)	(0.045)	(0.057)	(0.048)	(0.033)
Small, Large in Past	4.866***	4.636***	4.643***	4.944***	4.865***	4.450***
	(0.065)	(0.063)	(0.062)	(0.075)	(0.065)	(0.041)
Medium	7.503***	7.292***	7.315***	7.696***	7.495***	7.234***
	(0.057)	(0.055)	(0.055)	(0.075)	(0.057)	(0.050)
Medium, Large in Past	7.511***	7.237***	7.251***	7.585***	7.503***	7.122***
	(0.080)	(0.079)	(0.079)	(0.094)	(0.081)	(0.074)
Large	7.671***	7.446***	7.450***	7.724***	7.676***	7.436***
	(0.056)	(0.054)	(0.054)	(0.068)	(0.056)	(0.050)
Market Classification Variable						
Average Number of Plants		X				
Lagged Average Plants			X			
Before 1983 Average Plants				X		
Construction Employment					X	
Category 2		1.053***	1.118***	0.225***	0.132**	
		(0.036)	(0.032)	(0.062)	(0.049)	
Category 3		1.668***	1.836***	0.348***	0.199**	
		(0.050)	(0.047)	(0.058)	(0.061)	
Category 4		2.293***	2.424***	0.482***	0.169*	
		(0.063)	(0.062)	(0.061)	(0.082)	
Constant	−3.805***	−2.985***	−2.970***	−4.089***	−3.715***	
	(0.065)	(0.066)	(0.062)	(0.089)	(0.090)	
Observations	409,850	409,850	409,850	260,170	409,850	409,850
Markets	2029	2029	2029	2029	2029	2029
Log-Likelihood	−45,695	−44,483	−44,304	−27,334	−45,682	−39,670
χ^2	44,067	47,153	46,207	29,860	44,985	284,475

[a]Standard errors are clustered by market. *, **, *** indicate statistical significance at the 5%, 1%, and 0.1% levels, respectively.

using categories based on the number of firms before 1983 (d), or using information about the average level of construction demand (e), does not replicate market-fixed effect estimates, and, in fact, mimics not having any market controls whatsoever. Second, while it is true that using the average number of firms over time conditions on an endogenous variable, I can equally easily use the lagged number of firms that does not condition on an endogenous variable and obtain virtually identical results. Thus, the issue of endogeneity is of limited practical importance in the use of the average number of firms over time as a grouping.

C.2. *Alternative Market Categories From Market-Fixed Effects*

In this section, I present an alternative procedure for constructing market categories based on values of the market-fixed effect. Consider the binary logit model:

$$y_{imt} = 1(\alpha X_{imt} + \xi_m > \varepsilon_{imt}).$$

I can construct market categories based on estimates of the market fixed effect variable ε_{imt}, using the following procedure:

Step 1: Run a conditional logit (with market-fixed effects) on the number of active plants to recover parameters $\hat{\alpha}$, that is, everything except the market-fixed effect ξ^m. Note that we can get α without the problem of incidental parameters using a conditional logit.

Step 2: Create the variable $Z^{\hat{\alpha}} = \hat{\alpha}X$, the part of the covariates that is not the market-fixed effect.

Step 3: Run the logit on the model

$$y_{imt} = 1(Z^{\alpha}_{imt} + \xi_m > \varepsilon_{imt}),$$

which can be done separately for each market in the data. Note that this means that I am estimating market-fixed effects $\hat{\xi}^m$ (for which I need the number of time periods to be large).

Step 4: Use estimated $\hat{\xi}^m$ to form groups of markets.

The estimated $\hat{\xi}^m$ has the distribution in Table S.II.

TABLE S.II

DISTRIBUTION OF MARKET-FIXED EFFECT $\hat{\xi}^m$

Percentile	10	25	50	75	90
$\hat{\xi}^m$	−3.7	−3.0	−2.1	−1.4	−0.7

TABLE S.III

BINARY LOGIT REGRESSIONS OF THE DECISION TO HAVE AN ACTIVE PLANT WITH
MARKET-FIXED EFFECTS AND MARKET-CATEGORY EFFECTS

Dependent Variable: Activity	I	II (FE[†])	III (μ)	IV	V	VI	VII
Log County Construction	0.13	0.03	−0.03	0.053	0.051	0.030	0.024
Employment	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
First Competitor	−1.40	−2.02	−1.81	−2.14	−2.25	−2.26	−2.27
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Second Competitor	0.00	−0.64	−0.53	−0.56	−0.61	−0.61	−0.62
	(0.04)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Third Competitor	0.03	−0.40	−0.36	−0.26	−0.33	−0.34	−0.34
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Log Competitors Above 4	0.02	−0.19	−0.12	0.04	0.00	−0.13	−0.14
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Small	5.89	5.59	5.70	6.49	6.46	6.46	6.46
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Small, Medium in Past	5.67	5.22	5.39	6.33	6.30	6.29	6.30
	(0.05)	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Small, Large in Past	4.87	4.45	4.64	5.92	5.87	5.87	5.87
	(0.07)	(0.04)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Medium	7.50	7.23	7.30	7.34	7.34	7.34	7.34
	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Medium, Large in Past	7.51	7.12	7.24	7.29	7.26	7.25	7.25
	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Large	7.67	7.44	7.45	7.50	7.50	7.50	7.50
	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Market Classification Variable[◇]							
4 Fixed Effect Groups				X			
10 Fixed Effect Groups					X		
20 Fixed Effect Groups						X	
40 Fixed Effect Groups							X
Observations	409,850	409,850	409,850	405,143	405,143	405,143	405,143
Markets	2029	2029	2029	2014	2014	2014	2014
Log-Likelihood	−45,695	−44,483	−44,304	−35,000	−35,000	−35,000	−35,000
χ^2	44,067	47,153	46,207	55,000	55,000	61,000	81,000

[†]Market fixed-effects implemented via a conditional logit.

[◇]Market classification variable constructed using the procedure described in the text.

Table S.III shows the binary logit results on activity, where columns IV, V, VI, and VII show fixed effects constructed by rounding $\hat{\xi}^m$ into 4, 10, 20, and 40 categories (where each category contains the same number of markets). Notice that using these fixed effect categories yields similar results to Column I (fixed effects) and Column II (market categories). However, to match the market-category effects in Column III, in particular to get similar effects of log

competitors above 4, I need to have at least 10 market groups. In the estimates I will show you, I use 20 market categories of $\hat{\mu}^m$.

APPENDIX D: IDENTIFICATION OF FIXED COSTS, SUNK COSTS, AND SCRAP VALUES

To show the intuition behind the identification of fixed costs, sunk costs, and scrap values, I will use a simplified model. Suppose that I have variable profits $V(X)$ where X are time invariant profit shifters, fixed cost f, entry costs ψ, and exit costs ϕ. Then the entry and exit rules in a stationary environment are:

- Enter iff:

$$\sum_{t=0}^{\infty} \beta^t \big(V(X) - f\big) = \frac{V(X) - f}{1 - \beta} \geq \psi.$$

- Exit iff:

$$\sum_{t=0}^{\infty} \beta^t \big(V(X) - f\big) = \frac{V(X) - f}{1 - \beta} < \phi.$$

In this case it is clear that f, ψ, and ϕ are linearly independent. Adding future exit rates, δ (which at this point are generated by shocks to the exit value $\phi + \varepsilon$ that I do not want to put in this simple model), will adjust these equations to:

- Enter iff:

$$\frac{V(X) - f}{1 - \beta(1 - \delta)} + \frac{\phi}{1 - \beta\delta} \geq \psi.$$

- Exit iff:

$$\frac{V(X) - f}{1 - \beta(1 - \delta)} + \frac{\phi}{1 - \beta\delta} < \phi.$$

Again, we have the same collinearity problem as before. However, if future exit rates are different in different markets (say, due to differences in future demand shocks, such as a market at the top demand level, versus one at the bottom demand level), then we have a $\delta(X)$ that depends on the state X. This allows us to separately identify f and ϕ in the exit equation given that we know $V(X)$ and $\delta(X)$:

$$\frac{V(X) - f}{1 - \beta(1 - \delta(X))} + \frac{\phi}{1 - \beta\delta(X)} < \phi.$$

Now, given that we know $\hat{f}$ and $\hat{\phi}$, the entry equation becomes

$$\frac{V(X) - \hat{f}}{1 - \beta(1 - \delta(X))} + \frac{\hat{\phi}}{1 - \beta\delta(X)} \geq \psi.$$

So, formally, I can separately identify f, ϕ, and ψ. What makes this difficult is that I need enough variation in $\delta(X)$ for this to work, and this variation is not very important either in Monte Carlo simulations or in the data.

APPENDIX E: SIMULATED INDIRECT INFERENCE ESTIMATION

The simulated indirect inference estimator used in equation (S.19) uses the choice probabilities $\Psi(a|x, \Gamma, \theta)$ as an outcome vector, that is, $\tilde{\mathbf{y}}_n = \Psi(a|x, \Gamma, \theta)$. Typically, one would sample outcomes $\mathbf{y}_n$ from the choice probabilities $\Psi(a|x, \Gamma, \theta)$. I can show that using the $\tilde{\mathbf{y}}_n$ is equivalent to sampling actions as the number of actions tends to infinity.

Denote the outcome vectors $\mathbf{y}_n^s$ as

$$(\text{S.13}) \quad \mathbf{y}_n^s = \begin{bmatrix} 1(a_n^s = \text{small}) \\ 1(a_n^s = \text{medium}) \\ 1(a_n^s = \text{big}) \end{bmatrix},$$

where the action $a_n^s \sim \Psi(\cdot|x, \Gamma, \theta)$ is drawn from the choice probabilities Ψ. The simulation draws are indexed from $s = 1, \ldots, S$. The $\beta^S(\theta)$ coefficient is estimated using outcome vectors $\{y_n^s\}_{\{s=1,\ldots,S\}, n}$. The criterion function using S simulation draws of actions is thus

$$(\text{S.14}) \quad \mathcal{Q}^S(\theta) = (\hat{\beta} - \tilde{\beta}(\theta))' \mathbf{W}(\hat{\beta} - \tilde{\beta}(\theta)).$$

E.1. *Consistency Proof*

In this section, I show conditions under which the procedure I use in this paper is a consistent estimator of θ. Specifically, I show the conditions that need to be satisfied for Proposition 1 on p. S89 in Gourieroux, Monfort, and Renault (1993), dealing with the consistency of indirect inference estimators, to be satisfied.

Define the criterion function used to compute $\tilde{\beta}(\theta)$ (for a given value of θ) as

$$(\text{S.15}) \quad \mathcal{S}^{N,K}(\beta, \theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} \left[1(a_n^k = \text{small}) - Z_n \beta_s \right]^2$$
$$+ \left[1(a_n^k = \text{medium}) - Z_n \beta_m \right]^2$$
$$+ \left[1(a_n^k = \text{large}) - Z_n \beta_l \right]^2,$$

where N denotes the number of observations and K denotes the number of simulation draws to draw actions a_n^k from the policy function $\psi(a_n|x_n, \theta, \Gamma(\hat{P}^N, \hat{D}^N))$. Note that $\mathcal{S}^{N,K}(\beta, \theta)$ is the criterion used in OLS estimation, just the sum of squared errors.

The first step is to show that I can replace draws of a_n^k with the actual policy function ψ, or in other words, $\mathcal{S}^{N,K}(\beta, \theta) \to^{\text{a.s.}} \mathcal{S}^{N,\infty}(\beta, \theta)$ uniformly as $K \to \infty$.

THEOREM 1: *As the number of simulation draws K tends to infinity,* $\mathcal{S}^{N,K}(\beta, \theta) \to^{\text{a.s.}} \mathcal{S}^{N,\infty}(\beta, \theta)$ *uniformly.*

PROOF: I show the proof using only the choice to be small to lighten the notation, but the proof extends to as many actions as I want:

$$(S.16) \quad \mathcal{S}^{N,K}(\beta, \theta) = \sum_{n=1}^{N} \frac{1}{K} \sum_{k=1}^{K} \left[1(a_n^k = \text{small}) - Z_n \beta_s \right]^2$$

$$= \sum_{n=1}^{N} (Z_n \beta_s)^2 + \sum_{n=1}^{N} \sum_{k=1}^{K} \frac{1}{K} 1(a_n^k = \text{small})^2$$

$$- 2 \sum_{n=1}^{N} Z_n \beta_s \sum_{k=1}^{K} \frac{1}{K} 1(a_n^k = \text{small}).$$

As $K \to \infty$, $\sum_{k=1}^{K} \frac{1}{K} 1(a_n^k = \text{small}) \to \psi(a_n = \text{small}|x_n, \theta, \Gamma(\hat{P}, \hat{D}))$, since this is just an average, and $\sum_{k=1}^{K} \frac{1}{K} 1(a_n^k = \text{small})^2 \to \psi(a_n = \text{small}|x_n, \theta, \Gamma(\hat{P}, \hat{D}))^2$. Thus I can rewrite $\mathcal{S}^{N,\infty}(\beta, \theta)$ as

$$(S.17) \quad \mathcal{S}^{N,\infty}(\beta, \theta) = \sum_{n=1}^{N} (Z_n \beta_s)^2 + \sum_{n=1}^{N} \psi\left(a_n = \text{small}|x_n, \theta, \Gamma(\hat{P}, \hat{D})\right)^2$$

$$- 2 \sum_{n=1}^{N} Z_n \beta_s \psi\left(a_n = \text{small}|x_n, \theta, \Gamma(\hat{P}, \hat{D})\right)$$

$$= \sum_{n=1}^{N} \left[\psi\left(a_n = \text{small}|x_n, \theta, \Gamma(\hat{P}, \hat{D})\right) - Z_n \beta_s \right]^2.$$

Second, I need to show that $\mathcal{S}^{N,\infty}(\beta, \theta) \to^{\text{a.s.}} \mathcal{S}^{0,\infty}(\beta, \theta)$ as $N \to \infty$. The first condition is that the linear probability estimator is consistent, which is just an outcome of the OLS estimator being a consistent estimator, which is a standard proof. However, I am not using the true $\Gamma^0(P^0, D^0)$, but an estimate of $\Gamma(\hat{P}, \hat{D})$ due to sampling error in the conditional choice probabilities P and the demand transition process D, as well as approximation error in the computation of Γ. The CCP's $\hat{P}^N \to P^0$, which happens since I am using a consistent estimator of the CCP's, just a parametric multinomial logit, which is consistent using the usual proofs on the consistency of M-estimators. Likewise, $\hat{D}^N \to D^0$

as $N \to \infty$, since I am using a consistent estimator of D, just a bin estimator where the number of bins is fixed as $N \to \infty$. Now, the next point is to show that $\Gamma^L(P^0, D^0) \to \Gamma^0(P^0, D^0)$ as the number of iterations L in the DASA goes to infinity. It will be difficult to show convergence of the DASA, since to my knowledge there is no proof of the convergence of algorithms that compute the solutions to games (in contrast to single agent problems). However, the Fershtman and Pakes (2012) convergence criterion can be used to check the convergence of the DASA, and I can send the tolerance of the Fershtman–Pakes criterion to 0 as $N \to \infty$.[7]

The convergence of $\Gamma^L(P^0, D^0) \to \Gamma(P^0, D^0)$ implies the convergence of $\mathcal{S}^{N,K}(\beta, \theta) \to \mathcal{S}^{\infty,\infty}(\beta, \theta)$ as $K \to \infty$ and $N \to \infty$. This satisfies Assumption (A2) in indirect inference, where the notation (A2) refers to the numbering of assumptions in Gorieroux, Monfort, and Renault (1993).

Assumption (A3) of indirect inference requires that

$$(S.18) \quad \tilde{\beta}(\theta) = \underset{\beta}{\operatorname{argmax}} \, \mathcal{S}^{\infty,\infty}(\beta, \theta)$$

be a continuous function and have a unique value. Continuity is an outcome of the OLS structure of $\mathcal{S}$, while uniqueness occurs if Z_n is full rank and the dimension of β is smaller than the dimension of Z_n.

The final condition, (A4), requires that $\tilde{\beta}(\theta)$ be one to one and have full rank. I assume this condition, but notice that the dimension of β is larger than the dimension of θ, and I have checked that $\tilde{\beta}(\theta)$ is full rank in the estimation of the model.

Since conditions (A1), (A2), (A3), and (A4) are satisfied, then $\hat{\theta}$, defined as the minimizer of

$$(S.19) \quad \mathcal{Q}(\theta) = \big(\hat{\beta} - \tilde{\beta}(\theta)\big)' \mathbf{W} \big(\hat{\beta} - \tilde{\beta}(\theta)\big),$$

will be a consistent estimator of θ as $N \to \infty$. *Q.E.D.*

---

[7]Notice that, since there is a full support shock ε to the payoffs of any actions, Γ is computed correctly on the entire state space S, since the set of recurrent points is the entire state space, that is, $S = R$. The DASA used to compute Γ is a version of the Q-learning algorithm, where consistency proofs are provided for the single agent (non-game version) in Propositions 5.5 and 5.6 on pp. 248–249 in Bertsekas and Tsitsiklis (1996), which show conditions under which the DASA's (which is the game version of a Q-learning algorithm) computation of $\Gamma^{\cdot}$ converges with probability 1 to Γ^0. These conditions are (1) that policies are proper, that is, there is a positive probability that a firm will exit after t period, which is true in this context due to the full support of the shock distribution for each action, including the choice to exit; and (2) for improper policies, there is a negative infinite value of W for at least one state. Unfortunately, there is, to my knowledge, no proof that shows the convergence of the Q-learning algorithm in the context of a game.

APPENDIX F: Serial Correlation

The assumption that the unobserved state $\varepsilon_{a_i}^t$ are i.i.d. logits implies the following assumption:

Assumption 1—Serial Independence: *Unobserved states are serially independent, that is,* $\Pr(\varepsilon_i^t | \varepsilon_i^k) = \Pr(\varepsilon_i^t)$ *for* $k \neq t$.

Serial independence of unobserved components of a firm's profitability is violated by any form of persistent productivity difference between firms, or long-term reputations of ready-mix concrete operators. Note that, in the context of a dynamic game, unobserved states are a first-order problem since the size of the firm-level state x_i^t is severely restricted by the difficulty of keeping track of the joint distribution of the states of all firms.

I simulate the age profile of exit using the exit and size changes in Table II, which captures what the age profile of exit would look like in the absence of selection on an unobserved state. With a serially correlated unobserved state, as plants age their exit rate falls due to the effect of selecting out plants with a bad unobserved state. Figure S.I shows the exit haz-
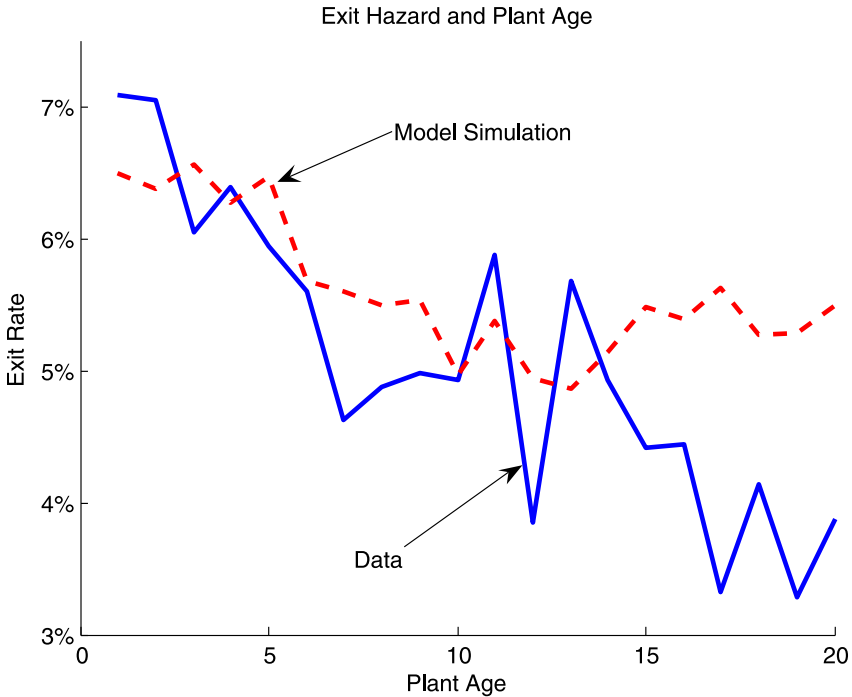

FIGURE S.I.—The data predict a slightly steeper decline of the exit hazard with age.

ard with age in the data and simulated data. Both the data and the simulation have the same average exit rate of about 6%, but the data have a somewhat steeper decline in exit rates over time, so a plant aged 20 years old has an exit rate of about 3.5% in the data, while the simulated data yield a exit rate of about 5.2%. This is consistent with most models of industry dynamics with a serially correlated unobserved state, and the active or passive learning models of Pakes and Ericson (1998) and Jovanovic (1982), but is a small effect compared to other industries such as restaurants, where we would worry more about unobserved states. I do not deal with serial correlation, and both the estimates and counterfactuals will be contaminated by this problem.

APPENDIX G: PRICE DATA

The Census Bureau does not generally collect price data. This job is left to the Bureau of Economic Analysis and the Bureau of Labor Statistics. However, following Syverson (2004a), we can generate prices using the following equation:

$$(S.20) \qquad p_{it}(c) = \frac{s_{it}(c)}{q_{it}(c)},$$

which is just sales of the commodity divided by quantity sold. While these "prices" may be good indicators of price dispersion (the application Syverson considered), they are particularly poor measures of actual plant prices, with an interquartile range over 2 log points (the third quartile is 100 times bigger than the first price quartile). This is probably because of how measurement error in the numerator and especially the denominator interact.

To reduce the impact of imputed data and measurement error on the dispersion of prices, I apply a version of Syverson's (2004b) procedures:

1. Hot Imputes in the data are identified as prices that satisfy the following:

$$(S.21) \qquad \left| p_i^t - p_j^t \right| < 0.0001 \quad \text{for some } i \text{ and } j \text{ in the data.}$$

I drop all prices that are hot imputes. Notice that this procedure will also eliminate cold imputes, defined as prices that equal the mode in the current year.

2. I trim the data by dropping observations that are less than 1/5 or more than 5 times the median price for the current year.

The deflated data are computed by $p_i^{Dt} = p_i^t / cpi^t$, where I normalize the cpi in 1977 to be equal to 1 (i.e., $cpi^t = \text{raw } cpi^t / \text{raw } cpi^{1977}$). This eliminates differences in price level across time, but does not incorporate differences in prices between regions.

APPENDIX H: ADDITIONAL TABLES AND FIGURES

TABLE S.IV

THE NUMBER OF BIRTHS, DEATHS, AND CONTINUERS IS FAIRLY STABLE OVER THE
LAST 25 YEARS

Year	Birth	Continuer	Death
1976	501	4737	N.A.
1977	557	4791	410
1978	327	5043	445
1979	392	5093	333
1980	271	5140	387
1981	313	5069	360
1982	313	4875	423
1983	273	4991	315
1984	328	4972	295
1985	309	4988	339
1986	300	5003	305
1987	390	4898	404
1988	270	5016	269
1989	248	4275	448
1990	194	4103	304
1991	220	3882	291
1992	214	4643	348
1993	133	3668	270
1994	163	3952	232
1995	196	3840	243
1996	195	3734	230
1997	338	4768	274
1998	239	4949	267
1999	320	4961	234

REFERENCES

BERTSEKAS, D., AND J. TSITSIKLIS (1996): *Neuro-Dynamic Programming*. Belmont, MA: Athena Scientific. [13]

FERSHTMAN, C., AND A. PAKES (2012): "Dynamics Games With Asymmetric Information: A Framework for Empirical Work," *Quarterly Journal of Economics*, 127 (4), 1611–1661. [2,5,13]

GOURIEROUX, C., A. MONFORT, AND E. RENAULT (1993): "Indirect inference," *Journal of Applied Econometrics*, 8, S85–S118. [11,13]

GOWRISANKARAN, G. (1999): "Efficient Representation of State Spaces for Some Dynamic Models," *Journal of Economic Dynamics and Control*, 23 (8), 1077–1098. [1]

JOVANOVIC, B. (1982): "Selection and the Evolution of Industry," *Econometrica*, 50 (3), 649–670. [15]

PAKES, A., AND R. ERICSON (1998): "Empirical Applications of Alternative Models of Firm and Industry Dynamics," *Journal of Economic Theory*, 79 (1), 1–45. [15]

PAKES, A., AND P. MCGUIRE (2001): "Stochastic Algorithms, Symmetric Markov Perfect Equilibrium, and the 'Curse' of Dimensionality," *Econometrica*, 69 (5), 1261–1281. [1]

POWELL, W. B. (2007): *Approximate Dynamic Programming: Solving the Curse of Dimensionality*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley. [2]

SYVERSON, C. (2004a): "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 112 (6), 1181–1222. [15]

———— (2004b): "Product Substitutability and Productivity Dispersion," *Review of Economics and Statistics*, 86 (2), 534–550. [15]

Stern School of Business, New York University, New York, NY 10012, U.S.A.; wexler@stern.nyu.edu.