Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Markov Decision Processes with Exogenous Variables

Robert L. Bray

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 65, No. 10, October 2019, pp. 4598–4606
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Markov Decision Processes with Exogenous Variables

Robert L. Bray[a]

[a] Kellogg School of Management, Northwestern University, Evanston, Illinois 60208-0814

Contact: r-bray@kellogg.northwestern.edu, http://orcid.org/0000-0003-2773-0663 (RLB)

Received: February 22, 2017
Revised: August 12, 2017; February 5, 2018; May 9, 2018
Accepted: June 20, 2018
Published Online in Articles in Advance: April 24, 2019

https://doi.org/10.1287/mnsc.2018.3158

Abstract. I present two algorithms for solving dynamic programs with exogenous variables: endogenous value iteration and endogenous policy iteration. These algorithms are always at least as fast as relative value iteration and relative policy iteration, and they are faster when the endogenous variables converge to their stationary distributions sooner than the exogenous variables.

History: Accepted by Yinyu Ye, optimization.

Keywords: Markov decision process • dynamic programming • endogenous value iteration • relative value iteration • exogenous variables

1. Introduction

An ergodic Markov decision process will eventually forget the current action as its state variables regress to their joint stationary distribution. After the variables reach stationarity, the action information is lost. All subsequent payoffs are independent of this action and have no bearing on it. Thus, not all future payoffs influence the current action—only those received before the action's memory has been purged. This idea underlies the relative value iteration and relative policy iteration algorithms of Morton (1971) and Morton and Wecker (1977), which disregard payoffs received after the state variables reach stationarity.

However, not all state variables must reach stationarity for the current action to be forgotten, because not all state variables encode action information. To be informative of an action, a variable must respond to the action. Because *exogenous* variables that evolve independently of the actions do not encrypt action information, the only payoffs that matter are those received before the *endogenous* variables reach stationarity.

My new solution algorithms—endogenous value iteration and endogenous policy iteration—exploit this insight. They iterate Bellman contractions until the endogenous state variables reach their joint stationary distribution, and then, they stop. These algorithms are faster than relative value and relative policy iteration when the exogenous variables are more persistent than the endogenous variables.

2. Markov Decision Process

An agent controls a Markov chain with a sequence of actions. In a given period, the agent observes exogenous state variable $x \in \mathbb{x}$ and endogenous state variable $y \in \mathbb{y}$ and chooses action $a \in \mathbb{a}$ in response. Action space $\mathbb{a}$ is compact, exogenous state space $\mathbb{x}$ has finite cardinality $X = |\mathbb{x}|$, endogenous state space $\mathbb{y}$ has finite cardinality $Y = |\mathbb{y}|$, and total state space $\mathbb{z} = \mathbb{x} \times \mathbb{y}$ has finite cardinality $Z = XY$. The state spaces are ordered: x_i, y_i, and z_i are the ith elements of $\mathbb{x}$, $\mathbb{y}$, and $\mathbb{z}$. The total state space $\mathbb{z}$ is arranged lexicographically:

$$\begin{array}{llll} z_1 = (x_1, y_1), & z_{Y+1} = (x_2, y_1), & \cdots & z_{(X-1)Y+1} = (x_X, y_1), \\ z_2 = (x_1, y_2), & z_{Y+2} = (x_2, y_2), & \cdots & z_{(X-1)Y+2} = (x_X, y_2), \\ \vdots & \vdots & & \vdots \\ z_Y = (x_1, y_Y), & z_{2Y} = (x_2, y_Y), & \cdots & z_{XY} = (x_X, y_Y). \end{array}$$

In general, $z_i = (x^i, y^i)$, where $x^i = x_{1+(i-\mathrm{mod}(i,Y))/Y}$ and $y^i = y_{1+\mathrm{mod}(i-1,Y)}$.

The agent's actions influence the endogenous variable but not the exogenous variable. Specifically, taking action a in state (x, y) sets the probability mass function of the next period's state variables to $f(x', y'|a, x, y) = f_x(x'|x)f_y(y'|a, x, y)$. Function f is continuous in a.[1]

The agent receives utility $u(a, x, y)$ from taking action a in state (x, y). The agent aims to maximize its expected infinite horizon utility under discount factor $\beta < 1$. Function u is bounded and continuous in a.

The following Bellman equation implicitly defines the agent's optimal value function:

$$v^*(x, y) = \max_{a \in \mathbb{a}} u(a, x, y) + \beta \sum_{x' \in \mathbb{x}} \sum_{y' \in \mathbb{y}} f(x', y'|a, x, y) v^*(x', y').$$

The corresponding optimal policy function is

$$p^*(x, y) = \arg\max_{a \in \mathbb{a}} u(a, x, y) + \beta \sum_{x' \in \mathbb{x}} \sum_{y' \in \mathbb{y}} f(x', y'|a, x, y) v^*(x', y').$$

Henceforth, v^* and p^* represent optimal value and policy functions, respectively, and v and p represent generic value and policy functions, respectively.

3. Variable Glossary

1. $\otimes$ is the Kronecker product symbol.
2. $\delta_i(j)$ is the length-i unit vector indicating the jth position.
3. $\delta(j)$ is the length-Z unit vector indicating the jth position.
4. ι_i is the length-i vector of ones.
5. $\iota = \iota_Z$ is the length-Z vector of ones.
6. I_i is the rank-i identity matrix.
7. $I = I_Z$ is the rank-Z identity matrix.
8. $\Delta_i = I_i - \iota_i\delta_i(1)'$ is the $i \times i$ differencing operator: the jth element of $\Delta_i\alpha$ is $\alpha_j - \alpha_1$.
9. $\Delta = \Delta_Z$ is the $Z \times Z$ differencing operator: the ith element of ΔV is $v(x^i, y^i) - v(x_1, y_1)$.
10. $\Xi = \iota_Y\iota_Y'/Y$ is the matrix of ones divided by Y.
11. $\Omega = I_Y - \Xi$ is the $Y \times Y$ demeaning operator: the ith element of $\Omega\alpha$ is $\alpha_i - \sum_{j=1}^{Y}\alpha_j/Y$.
12. $\Lambda = I - I_X \otimes \Xi = I_X \otimes \Omega$ is the $Z \times Z$ endogenizing operator: the ith element of ΛV is $v(x^i, y^i) - \sum_{j=1}^{Y} v(x^i, y_j)/Y$.
13. V is the value function, the length-Z vector with ith element $v(x^i, y^i)$.
14. $\breve{V} = \Delta V$ is the relative value function, the length-Z vector with ith element $v(x^i, y^i) - v(x_1, y_1)$.
15. $\bar{V} = \Lambda V$ is the endogenous value function, the length-Z vector with ith element $v(x^i, y^i) - \sum_{j=1}^{Y} \cdot v(x^i, y_j)/Y$.
16. $U(p)$ is the utility function, the length-Z vector with ith element $u(p(x^i, y^i), x^i, y^i)$.
17. F_x is the exogenous state transition matrix, the $X \times X$ matrix with ijth element $f_x(x_j|x_i)$.
18. $F_y(p, x)$ is the endogenous state transition matrix, the $Y \times Y$ matrix with ijth element $f_y(y_j|p(x,\ y_i), x, y_i)$.
19. $F(p) = \sum_{i=1}^{X}(\delta_X(i)\delta_X(i)'F_x) \otimes F_y(p, x_i)$ is the state transition matrix, the $Z \times Z$ matrix with ijth element $f(x^j, y^j|p(x^i, y^i), x^i, y^i)$.
20. $\mathbb{V} = \mathbb{R}^Z$ is the set of value vectors.
21. $\mathbb{V}_x = \text{colspace}(I_X \otimes \iota_Y)$ is the set of exogenous value vectors.
22. $\mathbb{V}_y = \mathbb{V}_x^{\perp}$ is the set of endogenous value vectors.
23. $\mathbb{p} = \mathbb{a}^{\mathbb{z}}$ is the set of policy functions.
24. γ is the Bellman contraction function: $\gamma(p, V) = U(p) + \beta F(p)V$.
25. π is the policy improvement operator: $\pi V = \arg\max_{p \in \mathbb{p}} \iota'\gamma(p, V)$.[2]
26. ν is the value iteration operator: $\nu V = \gamma(\pi V, V)$.
27. $\breve{\nu} = \Delta\nu$ is the relative value iteration operator.
28. $\bar{\nu} = \Lambda\nu$ is the endogenous value iteration operator.
29. η is the policy valuation operator: $\eta p = (I - \beta F(p))^{-1}U(p)$.
30. $\breve{\eta} = \Delta\eta$ is the relative policy valuation operator.
31. $\bar{\eta} = \Lambda\eta$ is the endogenous policy valuation operator.
32. η_t is the t-step policy valuation operator: $\eta_t p = \sum_{\tau=0}^{t}\beta^\tau F(p)^\tau U(p)$.
33. $\breve{\eta}_t = \Delta\eta_t$ is the relative t-step policy valuation operator.
34. $\bar{\eta}_t = \Lambda\eta_t$ is the endogenous t-step policy valuation operator.
35. $\rho = \pi\eta$ is the policy iteration operator.
36. $\breve{\rho} = \pi\breve{\eta}$ is the relative policy iteration operator.
37. $\bar{\rho} = \pi\bar{\eta}$ is the endogenous policy iteration operator.
38. $\theta_\epsilon = \epsilon(1-\beta)/(2\beta)$ is the algorithm stopping threshold.
39. ϕ is the function that maps a square matrix to its spectral radius (largest eigenvalue modulus).
40. σ is the function that maps a square matrix to its spectral subradius (second largest eigenvalue modulus).
41. O represents big O convergence: $f(t)$ is $O(\lambda^t)$ if there exist M and t_0, such that $|f(t)| < M|\lambda^t|$ for all $t > t_0$.
42. O^+ represents the "high order" convergence of Morton and Wecker (1977): $f(t)$ is $O^+(\lambda^t)$ if $f(t)$ is $O((\lambda + \epsilon)^t)$ for all $\epsilon > 0$.

4. Traditional Algorithms

The value iteration algorithm iteratively sets $V_t = \nu V_{t-1}$ until $\|V_t - V_{t-1}\| < \theta_\epsilon$, at which time policy πV_t is ϵ optimal. The policy iteration algorithm iteratively sets $p_t = \rho p_{t-1}$ until $\|\nu\eta p_{t-1} - \eta p_{t-1}\| < \theta_\epsilon$, at which time policy p_t is ϵ optimal. The most difficult part of policy iteration is calculating value $V_t = \eta p_{t-1}$. A common way to do so is to set $\eta_0 p = U(p)$ and iterate $\eta_t p = \gamma(p, \eta_{t-1}p)$ to convergence (note that $\lim_{t\to\infty}\eta_t p = \eta p$).

The relative value iteration algorithm of Morton and Wecker (1977) iteratively sets $\breve{V}_t = \breve{\nu}\breve{V}_{t-1}$ until $\|\breve{V}_t - \breve{V}_{t-1}\| < \theta_\epsilon$, at which time policy $\pi\breve{V}_t$ is ϵ optimal. The relative policy iteration algorithm of Morton (1971) iteratively sets $p_t = \breve{\rho}p_{t-1}$ until $\|\breve{\nu}\eta p_{t-1} - \eta p_{t-1}\| < \theta_\epsilon$, at which time policy p_t is ϵ optimal. The most difficult part of relative policy iteration is calculating relative value $\breve{V}_t = \breve{\eta}p_{t-1}$. One does so by setting $\breve{\eta}_0 p = \Delta U(p)$ and iterating $\breve{\eta}_t p = \Delta\gamma(p, \breve{\eta}_{t-1}p)$ to convergence (note that $\lim_{t\to\infty}\breve{\eta}_t p = \breve{\eta}p$).

The relative algorithms are faster than the standard algorithms when the Markov chain is ergodic. Although the standard algorithms consider all utilities with meaningful discounted expectation, the relative algorithms consider only the utilities received before the state variables revert back to their stationary distribution—the utilities received thereafter are moot, because the Markov chain has forgotten the current action by then. Disregarding these superfluous utilities expedites the computation. Morton (1971) and Morton and Wecker (1977) formalized this insight with the following "strong convergence" results.[3]

1. If the Markov chain is ergodic under policy p^*, then relative value iteration converges faster than value iteration: whereas $\|V^* - v^t 0\|$ is $O(\beta^t)$, $\|\check{V}^* - \check{v}^t 0\|$ is $O^+(\beta^t \sigma(F(p^*))^t)$, where $\sigma(F(p^*)) < 1$.

2. If the Markov chain is ergodic under policy p, then relative policy iteration's valuation step converges faster under policy p than policy iteration's valuation step: whereas $\|\eta p - \eta_t p\|$ is $O(\beta^t)$, $\|\check{\eta} p - \check{\eta}_t p\|$ is $O^+(\beta^t \sigma(F(p))^t)$, where $\sigma(F(p)) < 1$.

5. Endogenous Algorithms

Whereas relative value iteration and relative policy iteration disregard utilities incurred after all state variables reach stationarity, endogenous value iteration and endogenous policy iteration disregard utilities incurred after all endogenous state variables reach stationarity. The endogenous algorithms are never slower than the relative algorithms, and they are strictly faster when the endogenous variables converge to their limiting distribution faster than the exogenous variables. My algorithms exploit the following results.

Proposition 1.

1. *The space of feasible value functions is the direct sum of an exogenous space, which is the column space of $I_X \otimes \iota_Y$, and an endogenous space, which is the orthogonal complement of the exogenous space:*

$$\begin{aligned} &\mathbb{V} = \mathbb{V}_x \oplus \mathbb{V}_y, \\ \textit{where}\quad &\mathbb{V}_x = \text{colspace}(I_X \otimes \iota_Y) \\ \textit{and}\quad &\mathbb{V}_y = \mathbb{V}_x^{\perp}. \end{aligned}$$

2. *Endogenizing operator $\Lambda = I_X \otimes \Omega$ projects onto the endogenous space: if $V \in \mathbb{V}$, then $\Lambda V \in \mathbb{V}_y$ and $(I - \Lambda)V \in \mathbb{V}_x$.*

3. *The policy function only responds to the endogenous value function: $\pi V = \pi \Lambda V$.*

Proposition 2. *The optimal endogenous value function identifies the optimal value and policy functions: $V^* = \bar{V}^* + ((I_X - \beta F_x)^{-1} \otimes \Xi)\nu \bar{V}^*$ and $p^* = \pi \bar{V}^*$.*

Corollary 1. *The vector of ones is exogenous: $\iota \in \mathbb{V}_x$.*

Corollary 1 establishes that the policy function is invariant to uniform shifts in the value function: making all states \$1 more valuable does not affect the policy, because the agent receives the extra buck regardless of the action. Because it is moot, the relative algorithms disregard the portion of the value function attributable to ι (the span of ι is the null space of Δ). However, Propositions 1 and 2 indicate that we can push this idea further: rather than nullify the span of ι, we can nullify the entire $\mathbb{V}_x$ subspace.[4]

The endogenous algorithms do exactly that. Endogenous value iteration iteratively sets $\bar{V}_t = \bar{v}\bar{V}_{t-1}$ until $\|\bar{V}_t - \bar{V}_{t-1}\| < \theta_\epsilon$. Additionally, endogenous policy iteration iteratively sets $p_t = \bar{\rho} p_{t-1}$ until $\|\bar{v}\bar{\eta} p_{t-1} - \bar{\eta} p_{t-1}\| < \theta_\epsilon$. The most difficult part of endogenous policy iteration is calculating endogenous value $\bar{V}_t = \bar{\eta} p_{t-1}$. One does so by setting $\bar{\eta}_0 p = \Lambda U(p)$ and iterating $\bar{\eta}_t p = \Lambda \gamma(p, \bar{\eta}_{t-1} p)$ to convergence (note that $\lim_{t\to\infty} \bar{\eta}_t p = \bar{\eta} p$).

The following results establish that the endogenous algorithms weakly dominate the relative algorithms.

Proposition 3. *The endogenous value iteration and endogenous policy iteration algorithms yield ϵ-optimal policies.*

Proposition 4.

1. *Endogenous value iteration converges at least as fast as relative value iteration: whereas $\|\check{V}^* - \check{v}^t 0\|$ is $O^+(\beta^t \sigma(F(p^*))^t)$, $\|\bar{V}^* - \bar{v}^t 0\|$ is $O^+(\beta^t \phi(\Lambda F(p^*))^t)$, where $\phi(\Lambda F(p^*)) \le \sigma(F(p^*))$.*

2. *Endogenous policy iteration's valuation step converges at least as fast as relative policy iteration's valuation step: whereas $\|\check{\eta} p - \check{\eta}_t p\|$ is $O^+(\beta^t \sigma(F(p))^t)$, $\|\bar{\eta} p - \bar{\eta}_t p\|$ is $O^+(\beta^t \phi(\Lambda F(p))^t)$, where $\phi(\Lambda F(p)) \le \sigma(F(p))$.*

Proposition 5.

1. *Endogenous value iteration converges faster than relative value iteration when $\sigma(F_x)$ exceeds $\max_{x\in\mathbb{x}} \|\Omega F_y(p^*, x)\|$ for some matrix norm $\|\cdot\|$.*

2. *Under policy p, endogenous policy iteration's valuation step converges faster than relative policy iteration's valuation steps when $\sigma(F_x)$ exceeds $\max_{x\in\mathbb{x}} \|\Omega F_y(p, x)\|$ for some matrix norm $\|\cdot\|$.*

Corollary 2.

1. *Endogenous value iteration converges faster than relative value iteration when $\sigma(F_x)$ exceeds*
   (a) *the Euclidean norm of vec$(\Omega F_y(p^*, x))$ for all $x \in \mathbb{x}$,*
   (b) *the largest singular value of $\Omega F_y(p^*, x)$ for all $x \in \mathbb{x}$,*
   (c) *the maximum absolute row sum of $\Omega F_y(p^*, x)$ for all $x \in \mathbb{x}$,*
   (d) *the maximum absolute column sum of $\Omega F_y(p^*, x)$ for all $x \in \mathbb{x}$, or*
   (e) *the Hajnal matrix seminorm of $F_y(p^*, x)$ for all $x \in \mathbb{x}$.*

2. *Under policy p, endogenous policy iteration's valuation step converges faster than relative policy iteration's valuation steps when $\sigma(F_x)$ exceeds*
   (a) *the Euclidean norm of vec$(\Omega F_y(p, x))$ for all $x \in \mathbb{x}$,*
   (b) *the largest singular value of $\Omega F_y(p, x)$ for all $x \in \mathbb{x}$,*
   (c) *the maximum absolute row sum of $\Omega F_y(p, x)$ for all $x \in \mathbb{x}$,*
   (d) *the maximum absolute column sum of $\Omega F_y(p, x)$ for all $x \in \mathbb{x}$, or*
   (e) *the Hajnal matrix seminorm of $F_y(p, x)$ for all $x \in \mathbb{x}$.*

The Hajnal seminorm of a stochastic matrix is less than one when the matrix is *scrambling*: that is, when no two of its rows are orthogonal or when each pair of states can transition to a common third state in one period (Hajnal 1957). With this, Corollary 2 implies the following.

Corollary 3.

1. *Endogenous value iteration converges faster than relative value iteration when F_x is not ergodic and $F_y(p^*, x)$ is scrambling for each $x \in \mathbb{x}$.*
2. *Under policy p, endogenous policy iteration's valuation step converges faster than relative policy iteration's valuation step when F_x is not ergodic and $F_y(p, x)$ is scrambling for each $x \in \mathbb{x}$.*

The conditions of Corollary 3 imply that y is ergodic and that x is not.[5] In this case, the relative algorithms must consider all utilities not discounted to irrelevance, but the endogenous algorithms must consider only the subset of utilities incurred before y reaches stationarity. For example, a seasonal state variable that cycles between {winter, spring, summer, fall} would void relative value iteration's strong convergence but not endogenous value iteration's strong convergence as long as $F_y(p, x)$ scrambles.[6]

This scrambling assumption is unnecessary when y has constant-state transition matrix $F_y(p)$. In this case, the endogenous algorithms are faster when F_x's spectral subradius exceeds $F_y(p)$'s spectral subradius.

Corollary 4.

1. *Endogenous value iteration converges faster than relative value iteration when $F_y(p^*, x) = F_y(p^*)$ for all $x \in \mathbb{x}$ and $\sigma(F_x) > \sigma(F_y(p^*))$.*
2. *Under policy p, endogenous policy iteration's valuation step converges faster than relative policy iteration's valuation step when $F_y(p, x) = F_y(p)$ for all $x \in \mathbb{x}$ and $\sigma(F_x) > \sigma(F_y(p))$.*

The deterministic equivalence problem of Higle et al. (1990) and the quasi-open loop problem of Adelman and Mancini (2016) both satisfy the $F_y(p^*, x) = F_y(p^*)$ and $F_y(p, x) = F_y(p)$ conditions of Corollary 4.

6. Illustration

I now show endogenous value iteration with the market entry problem of Aguirregabiria and Magesan (2018). It was this problem and the ingenious Euler equations-based solution of Aguirregabiria and Magesan (2013) and Aguirregabiria and Magesan (2018) that gave me the idea for endogenous value iteration.

The action is a Boolean that indicates whether the firm participates in the market in year t ($a_t = 1$) or not ($a_t = 0$). The endogenous state variable is a Boolean that indicates the previous year's market participation: $y_t = a_{t-1}$. The exogenous state variable is a vector of length five: $x_t = [x_t^1, \cdots, x_t^5]' \in \mathbb{R}^5$, where x_t^1 is a firm productivity factor, x_t^2 and x_t^3 are variable profit factors, x_t^4 is a fixed cost factor, and x_t^5 is a market entry cost factor. Each of these five factors can take five values, and therefore, $X = 5^5 = 3,125$.

The exogenous factors evolve independently of one another. The ith exogenous factor follows Tauchen's (1986) finite-value approximation of the autoregressive process $x_t^i = \alpha_0^i + \alpha_1^i x_{t-1}^i + e_t^i$, where e_t^i is a standard normal, $\alpha_0^i = 0.2\mathbb{1}(i=1)$, and $\alpha_1^i = 0.9\mathbb{1}(i=1) + 0.6\mathbb{1}(i \neq 1)$.[7] Note that productivity shock x_t^1 is more persistent than the other exogenous factors (this will be important).

In period t, the agent receives utility $u(a_t, x_t, y_t) + e_t(a)$, where

$$u(a_t, x_t, y_t) = \underbrace{a_t}_{\text{activity}} \left(\underbrace{\exp(x_t^1)}_{\text{productivity}} \underbrace{(0.5 + x_t^2 - x_t^3)}_{\text{variable profit}} - \underbrace{(0.5 + x_t^4)}_{\text{fixed cost}} - \underbrace{(1 - y_t)}_{\text{prior inactivity}} \underbrace{(1 + x_t^5)}_{\text{entry cost}} \right),$$

and $e_t(0)$ and $e_t(1)$ are independent Gumbel random variables with mean zero. Following convention, I integrate over these error terms to express the value function and policy function in terms of x and y (Aguirregabiria and Mira 2010):

$$\begin{aligned} v^*(x, y) &= \iint \Big(\max_{a \in \{0,1\}} u(a, x, y) + e(a) \\ &\quad + \beta \sum_{x' \in \mathbb{x}} f(x'|x) v^*(x', a) \Big) de(0) de(1) \\ &= u(1, x, y) + \beta \sum_{x' \in \mathbb{x}} f(x'|x) v^*(x', 1) - \ln(p^*(x, y)) \end{aligned}$$

where

$$p^*(x, y) = \frac{\exp\Big(u(1, x, y) + \beta \sum_{x' \in \mathbb{x}} f(x'|x) v^*(x', 1)\Big)}{\sum_{a \in \{0,1\}} \exp\Big(u(a, x, y) + \beta \sum_{x' \in \mathbb{x}} f(x'|x) v^*(x', a)\Big)},$$

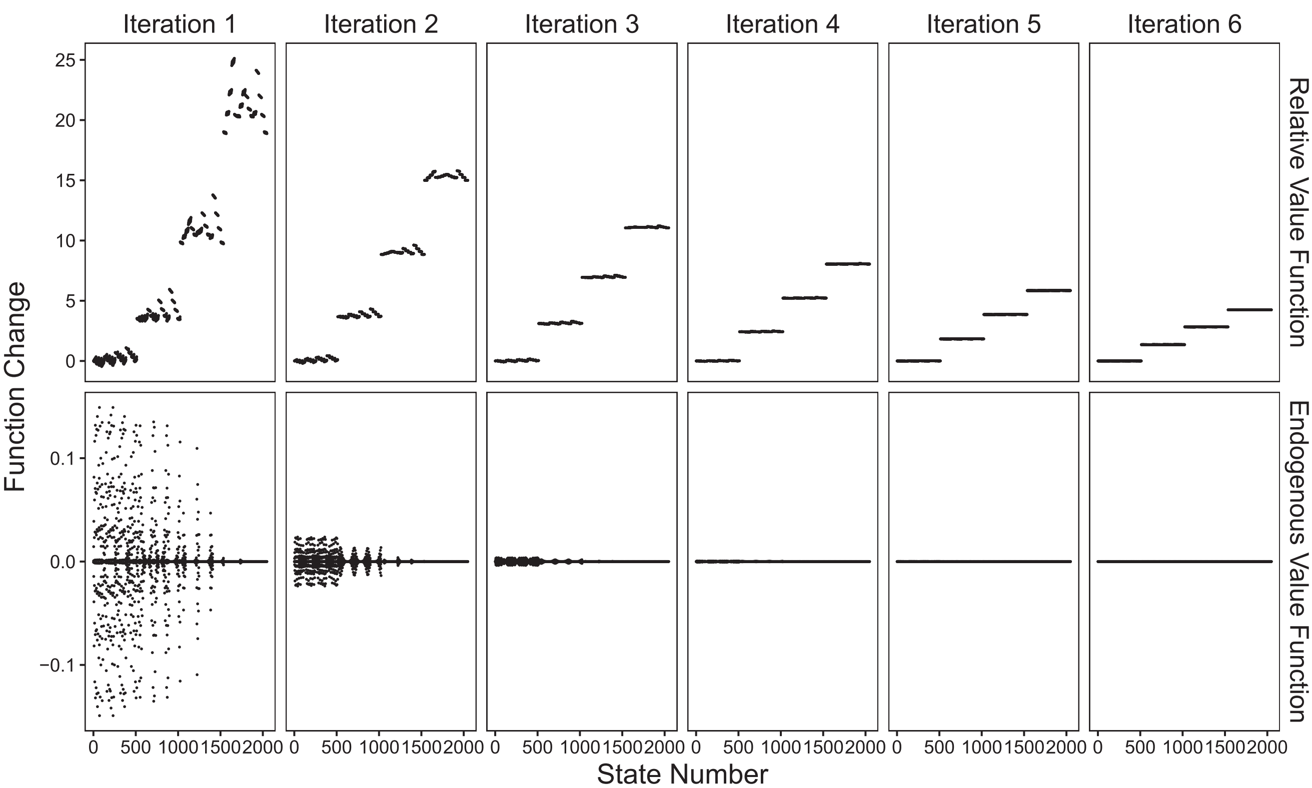
and $\quad \beta = 0.95.$

Note that the firm's policy function denotes the probability that it enters the market, conditional on x and y but not on $e(0)$ and $e(1)$.

I solve this problem with both endogenous value iteration and relative value iteration. Relative value iteration requires 62 Bellman contractions to converge to within a 10^{-6} tolerance, but endogenous value iteration requires only 10. The endogenous algorithm requires a sixth as many iterations, because it converges $\sigma(F(p^*))/\phi(\Lambda F(p^*)) = 0.56/0.082 = 6.52$ times as fast.[8]

Figure 1 depicts how the endogenous value function changes across the first six endogenous value iteration steps and how the relative value function changes across the first six relative value iteration steps. The relative value differences are more persistent, because they converge to a stair-step pattern that decays slowly. The staircase's five steps correspond to the five values of productivity factor x_t^1 (the most persistent exogenous variable). However, these stair-step value function changes do not influence the policy function, because the firm cannot influence its productivity.

Figure 1. Convergence of Relative and Endogenous Value Iteration



7. Conclusion

Not all of the value function constitutes useful information. Only the portion that influences the policy function is signal—the rest is noise. To filter out this noise, the relative value and policy iteration algorithms use projection matrix Δ, and the endogenous value and policy iteration algorithms use projection matrix Λ. The latter operator removes more noise, because it has a larger null space: whereas ΔV has $XY-1$ degrees of freedom, ΛV has only $X(Y-1)$ degrees of freedom.

The Λ operator projects away exogenous space $\mathbb{V}_x$. For simplicity, I have defined $\mathbb{V}_x = \text{colspace}(I_X \otimes \iota_Y)$. However, technically, I could expand the exogenous space to

$$\begin{aligned}\widetilde{\mathbb{V}}_x &= \{v \in \mathbb{V} \mid \pi(V+v) = \pi V \ \forall\ V \in \mathbb{V}\}\\ &= \{v \in \mathbb{V} \mid F(p_1)v = F(p_2)v \ \forall\ p_1, p_2 \in \text{image}(\pi)\}.\end{aligned}$$

Replacing Λ with $\widetilde{\Lambda}$, the projection matrix onto $\widetilde{\mathbb{V}}_y = \widetilde{\mathbb{V}}_x^{\perp}$ would yield even faster algorithms.

Proofs.

Lemma 1. *My proofs use the following identities:*

1. $\Delta_i \iota_i = 0$,
2. $\Xi^2 = \Xi$,
3. $\Delta^2 = \Delta$,
4. $\Lambda^2 = \Lambda$,
5. $\Lambda\Delta = \Lambda$,
6. $\Omega\Delta_Y = \Omega$,
7. $\Delta F(p)\Delta = \Delta F(p)$,
8. $\Lambda F(p)\Lambda = \Lambda F(p)$,
9. $\Omega F_y(p,x)\Omega = \Omega F_y(p,x)$,
10. $F(p)(M \otimes \Xi) = (F_x M) \otimes \Xi$ *for any* $X \times X$ *matrix* M,
11. $(M \otimes \Xi)\Lambda = \Lambda(M \otimes \Xi) = 0$ *for any* $X \times X$ *matrix* M, *and*
12. $\Delta(M \otimes \Xi) = (\Delta_X M) \otimes \Xi$ *for any* $X \times X$ *matrix* M.

Proof. Basic algebra yields these results. □

Proof of Proposition 1. The first two points are straightforward. The third point stems from Lemma 1.10, which implies that

$$\begin{aligned}\pi\Lambda V &= \arg\max_{p \in \mathbb{p}} \iota'(U(p) + \beta F(p)\Lambda V)\\ &= \arg\max_{p \in \mathbb{p}} \iota'(U(p) + \beta F(p)(I - I_X \otimes \Xi)V)\\ &= \arg\max_{p \in \mathbb{p}} \iota'(U(p) + \beta F(p)V - (F_x \otimes \Xi)V\\ &= \arg\max_{p \in \mathbb{p}} \iota'(U(p) + \beta F(p)V\\ &= \pi V. \quad \square\end{aligned}$$

Lemma 2. $\Lambda\gamma(p, \Lambda V) = \Lambda\gamma(p, V)$.

Proof. This follows from Lemma 1.8. □

Lemma 3. $\bar{\nu}^t V = \bar{\nu}\nu^{t-1}V = \Lambda\nu^t V$ *for any* $t \in \mathbb{N}_+$.

Proof. This follows from Lemma 2. □

Proof of Proposition 2. This proof has nine parts.

1. Proposition 1 implies that $\pi\bar{V}^* = \pi\Lambda V^* = \pi V^* = p^*$.
2. Proposition 1 and Lemma 3 imply that $\pi\nu^\tau\bar{V}^* = \pi\Lambda\nu^\tau\bar{V}^* = \pi\bar{\nu}^\tau\bar{V}^* = \pi\bar{V}^* = p^*$ for all $\tau\in\mathbb{N}$. This implies that $\nu^{t+1}\bar{V}^* - \nu^t\bar{V}^* = \beta F(p^*)(\nu^t\bar{V}^* - \nu^{t-1}\bar{V}^*)$ for all $t\in\mathbb{N}_+$. By induction, this implies that $\nu^{t+1}\bar{V}^* - \nu^t\bar{V}^* = \beta^t F(p^*)^t(\nu\bar{V}^* - \bar{V}^*)$ for all $t\in\mathbb{N}$.
3. Lemma 1.10 implies that $F(p)^t\,(F_x^s\otimes\Xi) = F(p)^{t-1}\cdot F(p)\,(F_x^s\otimes\Xi) = F(p)^{t-1}\,(F_x^{s+1}\otimes\Xi)$, which by induction, implies that $F(p)^t(I_X\otimes\Xi) = F(p)^t\,(F_x^0\otimes\Xi) = F(p)^0\cdot(F_x^t\otimes\Xi) = F_x^t\otimes\Xi$.
4. Lemma 1.2 implies that $F_x^t\otimes\Xi = (F_x^{t-1}\otimes\Xi)\cdot(F_x\otimes\Xi)$, which by induction, implies that $F_x^t\otimes\Xi = (F_x\otimes\Xi)^t$.
5. Points 3 and 4 imply that $F(p)^t(I-\Lambda) = (F_x\otimes\Xi)^t$ for any $t\in\mathbb{N}_+$.
6. Lemmas 1.2 and 1.11 imply that

$$\begin{aligned}(I-\beta F_x\otimes\Xi)&(\Lambda+(I_X-\beta F_x)^{-1}\otimes\Xi)\\ &=\Lambda+(I_X-\beta F_x)^{-1}\otimes\Xi-\beta F_x(I_X-\beta F_x)^{-1}\otimes\Xi\\ &=\Lambda+(I_X-\beta F_x)(I_X-\beta F_x)^{-1}\otimes\Xi\\ &=\Lambda+I_X\otimes\Xi\\ &=I.\end{aligned}$$

7. Point 6 implies that $(I-\beta F_x\otimes\Xi)^{-1} = \Lambda + (I_X-\beta F_x)^{-1}\otimes\Xi$.
8. Points 2 and 5 imply that

$$\begin{aligned}\nu^{t+1}\bar{V}^* - \nu^t\bar{V}^* &= \beta^t F(p^*)^t(\nu\bar{V}^*-\bar{V}^*)\\ &=\beta^t F(p^*)^t(\nu\bar{V}^*-\bar{\nu}\bar{V}^*)\\ &=\beta^t F(p^*)^t(\nu\bar{V}^*-\Lambda\nu\bar{V}^*)\\ &=\beta^t F(p^*)^t(I-\Lambda)\nu\bar{V}^*\\ &=\beta^t\,(F_x\otimes\Xi)^t\nu\bar{V}^*.\end{aligned}$$

9. Points 7 and 8 imply that

$$\begin{aligned}V^* &= \lim_{t\to\infty}\nu^t\bar{V}^*\\ &=\nu\bar{V}^* + \sum_{t=1}^{\infty}(\nu^{t+1}\bar{V}^*-\nu^t\bar{V}^*)\\ &=\beta^0(F_x\otimes\Xi)^0\nu\bar{V}^* + \sum_{t=1}^{\infty}\beta^t(F_x\otimes\Xi)^t\nu\bar{V}^*\\ &=(I-\beta F_x\otimes\Xi)^{-1}\nu\bar{V}^*\\ &=(\Lambda+(I_X-\beta F_x)^{-1}\otimes\Xi)\nu\bar{V}^*\\ &=\bar{V}^*+((I_X-\beta F_x)^{-1}\otimes\Xi)\nu\bar{V}^*.\end{aligned}$$ □

Proof of Corollary 1. This follows from Proposition 1. □

Proof of Proposition 3. Lemma 3 implies that $\lim_{t\to\infty}\bar{\nu}^t V = \lim_{t\to\infty}\Lambda\nu^t V = \Lambda V^* = \bar{V}^*$, and Proposition 1 implies that $\lim_{t\to\infty}\bar{\rho}^t p = \lim_{t\to\infty}(\pi\Lambda\eta)^t\eta p = \lim_{t\to\infty}(\pi\eta)^t p = \lim_{t\to\infty}\rho^t p = p^*$. These results imply that terminal conditions $\|\bar{V}_t-\bar{V}_{t-1}\|<\theta_\epsilon$ and $\|\bar{\nu}\bar{\eta}p_{t-1}-\bar{\eta}p_{t-1}\|<\theta_\epsilon$ hold after a finite number of iterations.

Now I show that the policy is ϵ optimal when these terminal conditions are first met. I focus on endogenous value iteration, but an equivalent argument holds for endogenous policy iteration. The proof has 10 parts.

1. Suppose that endogenous value iteration returns policy $\pi\bar{V}_t$ after terminal condition $\|\bar{V}_t-\bar{V}_{t-1}\|<\theta_\epsilon$, and define $V_0 = \bar{V}_{t-1} + ((I_X-\beta F_x)^{-1}\otimes\Xi)(\nu\bar{V}_{t-1}-\bar{V}_{t-1})$.
2. Lemma 1.11 implies that

$$\Lambda((I_X-\beta F_x)^{-1}\otimes\Xi)=0,$$
$$\Lambda\Big((\beta F_x(I_X-\beta F_x)^{-1})\otimes\Xi\Big)=0,$$
$$(I-\Lambda)((I_X-\beta F_x)^{-1}\otimes\Xi)=((I_X-\beta F_x)^{-1}\otimes\Xi),$$

and

$$(I-\Lambda)\Big((\beta F_x(I_X-\beta F_x)^{-1})\otimes\Xi\Big)=\Big((\beta F_x(I_X-\beta F_x)^{-1})\otimes\Xi\Big).$$

3. Point 2 implies that

$$\begin{aligned}\Lambda V_0 &= \Lambda\bar{V}_{t-1}+\Lambda((I_X-\beta F_x)^{-1}\otimes\Xi)(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\\ &=\Lambda\bar{V}_{t-1}.\end{aligned}$$

4. Proposition 1 and Point 3 imply that $\pi V_0 = \pi\Lambda V_0 = \pi\Lambda\bar{V}_{t-1} = \pi\bar{V}_{t-1}$.
5. Lemma 1.10 and Point 4 imply that

$$\begin{aligned}\nu V_0 &= U(\pi V_0)+\beta F(\pi V_0)V_0\\ &=U(\pi\bar{V}_{t-1})+\beta F(\pi\bar{V}_{t-1})\\ &\quad\cdot\Big(\bar{V}_{t-1}+((I_X-\beta F_x)^{-1}\otimes\Xi)(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\Big)\\ &=\nu\bar{V}_{t-1}+\beta F(\pi\bar{V}_{t-1})((I_X-\beta F_x)^{-1}\otimes\Xi)\\ &\quad\cdot(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\\ &=\nu\bar{V}_{t-1}+\big((\beta F_x(I_X-\beta F_x)^{-1})\otimes\Xi\big)(\nu\bar{V}_{t-1}-\bar{V}_{t-1}).\end{aligned}$$

6. Points 2 and 5 imply that

$$\begin{aligned}\Lambda\nu V_0 &= \Lambda\nu\bar{V}_{t-1}+\Lambda((\beta F_x(I_X-\beta F_x)^{-1})\otimes\Xi)\\ &\quad\cdot(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\\ &=\Lambda\nu\bar{V}_{t-1}.\end{aligned}$$

7. Points 2 and 5 imply that

$$\begin{aligned}(I-\Lambda)&(\nu V_0 - V_0)\\ &=\Big((I-\Lambda)\nu\bar{V}_{t-1}+((\beta F_x(I_X-\beta F_x)^{-1})\otimes\Xi)\\ &\quad\cdot(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\Big)\\ &\quad-\Big((I-\Lambda)\bar{V}_{t-1}+((I_X-\beta F_x)^{-1}\otimes\Xi)\\ &\quad\cdot(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\Big)\\ &=(I-\Lambda)(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\\ &\quad-(((I_X-\beta F_x)(I_X-\beta F_x)^{-1})\otimes\Xi)(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\\ &=(I-\Lambda)(\nu\bar{V}_{t-1}-\bar{V}_{t-1})-(I_X\otimes\Xi)(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\\ &=(I-\Lambda)(\nu\bar{V}_{t-1}-\bar{V}_{t-1})-(I-\Lambda)(\nu\bar{V}_{t-1}-\bar{V}_{t-1})\\ &=0.\end{aligned}$$

8. Points 3, 6, and 7 imply that

$$\begin{aligned}\nu V_0 - V_0 &= \Lambda(\nu V_0 - V_0) + (I-\Lambda)(\nu V_0 - V_0)\\ &= \Lambda(\nu V_0 - V_0)\\ &= \Lambda\nu \bar{V}_{t-1} - \Lambda \bar{V}_{t-1}\\ &= \bar{V}_t - \bar{V}_{t-1}.\end{aligned}$$

9. Points 1 and 8 imply that $\|\nu V_0 - V_0\| < \theta_\epsilon$.

10. Points 6 and 9 imply that standard value iteration returns policy $\pi\nu V_0 = \pi\Lambda\nu V_0 = \pi\Lambda\nu\bar{V}_{t-1} = \pi\bar{\nu}\bar{V}_{t-1} = \pi\bar{V}_t$ when initialized with starting value V_0. Because standard value iteration always returns an ϵ-optimal policy, $\pi\bar{V}_t$ must be ϵ optimal. □

Lemma 4. *Every eigenvalue of $\Delta_X F_x$ is an eigenvalue of $\Delta F(p)$.*

Proof. If $\Delta_X F_x v = \lambda v$, for some eigenvector v, then

$$\begin{aligned}\Delta F(p)(v\otimes\iota_Y) &= \Delta\sum_{i=1}^{X}\Big((\delta_X(i)\delta_X(i)'F_x)\otimes F_y(p,x_i)\Big)(v\otimes\iota_Y)\\ &= \Delta\sum_{i=1}^{X}(\delta_X(i)\delta_X(i)'F_x v)\otimes(F_y(p,x_i)\iota_Y)\\ &= \Delta\big((F_x v)\otimes\iota_Y\big)\\ &= (F_x v)\otimes\iota_Y - (\iota\delta(1)')((F_x v)\otimes\iota_Y)\\ &= (F_x v)\otimes\iota_Y - \Big((\iota_X\delta_X(1)')\otimes(\iota_Y\delta_Y(1)')\Big)\\ &\quad\cdot((F_x v)\otimes\iota_Y)\\ &= (F_x v)\otimes\iota_Y - (\iota_X\delta_X(1)'F_x v)\otimes\iota_Y\\ &= (\Delta_X F_x v)\otimes\iota_Y\\ &= \lambda v\otimes\iota_Y. \quad □\end{aligned}$$

Lemma 5. *Every eigenvalue of $\Lambda F(p)$ is also an eigenvalue of $\Delta F(p)$.*

Proof. The proof has nine parts.

1. Assume that there exists λ that is not an eigenvalue of $\Delta F(p)$ but is an eigenvalue of $\Lambda F(p)$, with corresponding eigenvector v.

2. Lemma 4 implies that λ is not an eigenvalue of $\Delta_X F_x$.

3. Point 2 implies that $\lambda \neq 0$, because $\Delta_X F_x \iota_X = \Delta_X \iota_X = 0\iota_X$.

4. Lemma 1.4 and Point 3 imply that $v = \Lambda F(p)v/\lambda = \Lambda(\Lambda F(p)v/\lambda) = \Lambda v$.

5. λ not being an eigenvalue of $\Delta F(p)$ implies that $\lambda I - \Delta F(p)$ is invertible.

6. Lemma 1.5 and Point 5 imply that

$$\begin{aligned}\Delta F(p)&\Big(v + (\lambda I - \Delta F(p))^{-1}(I-\Lambda)\Delta F(p)v\Big)\\ &= \Lambda\Delta F(p)v + (I-\Lambda)\Delta F(p)v\\ &\quad + \Delta F(p)(\lambda I - \Delta F(p))^{-1}(I-\Lambda)\Delta F(p)v\\ &= \Lambda F(p)v + \Big((\lambda I - \Delta F(p)) + \Delta F(p)\Big)\\ &\quad\cdot(\lambda I - \Delta F(p))^{-1}(I-\Lambda)\Delta F(p)v\\ &= \lambda\Big(v + (\lambda I - \Delta F(p))^{-1}(I-\Lambda)\Delta F(p)v\Big),\end{aligned}$$

which implies that $v + (\lambda I - \Delta F(p))^{-1}(I-\Lambda)\Delta F(p)v = 0$, because otherwise, λ would be an eigenvalue of $\Delta F(p)$.

7. Lemmas 1.10 and 1.12 imply that

$$\begin{aligned}(\lambda I &- \Delta F(p))((\lambda I_X - \Delta_X F_x)^{-1}\otimes\Xi)\\ &= \lambda(\lambda I_X - \Delta_X F_x)^{-1}\otimes\Xi - \Delta\Big((F_x(\lambda I_X - \Delta_X F_x)^{-1})\otimes\Xi\Big)\\ &= \lambda(\lambda I_X - \Delta_X F_x)^{-1}\otimes\Xi - (\Delta_X F_x(\lambda I_X - \Delta_X F_x)^{-1})\otimes\Xi\\ &= ((\lambda I_X - \Delta_X F_x)(\lambda I_X - \Delta_X F_x)^{-1})\otimes\Xi\\ &= I_X\otimes\Xi\\ &= I-\Lambda.\end{aligned}$$

8. Points 5 and 7 imply that $(\lambda I - \Delta F(p))^{-1}(I-\Lambda) = (\lambda I_X - \Delta_X F_x)^{-1}\otimes\Xi$.

9. Lemma 1.11 and Points 4, 6, and 8 imply that

$$\begin{aligned}0 &= \Lambda 0\\ &= \Lambda\Big(v + (\lambda I - \Delta F(p))^{-1}(I-\Lambda)\Delta F(p)v\Big)\\ &= \Lambda v + \Lambda((\lambda I_X - \Delta_X F_x)^{-1}\otimes\Xi)\Delta F(p)v\\ &= \Lambda v\\ &= v,\end{aligned}$$

which is a contradiction, because an eigenvector cannot be zero. Thus, the assumption in Point 1 must be incorrect. □

Lemma 6. *$\sigma(F) = \phi(\Delta_i F)$ for any $i\times i$ stochastic matrix F.*

Proof. I will first show that $\phi(\Delta_i F) \geq \sigma(F)$. Let λ be an eigenvalue of F with eigenvector v. Lemma 1.7 implies that $\Delta_i F\Delta_i v = \Delta_i F v = \lambda\Delta_i v$. Thus, λ is an eigenvector of $\Delta_i F$ when $\Delta_i v$ is nonzero, $\Delta_i v$ is nonzero when v is not in the null space of Δ_i, and v is not in the null space of Δ_i when it is not in the span of ι_i. Because ι_i is the eigenvector associated with F's largest eigenvalue (Puterman 2005, p. 595), the second largest eigenvalue of F is an eigenvalue of $\Delta_i F$.

I will now show that $\sigma(F) \geq \phi(\Delta_i F)$. Let λ be an eigenvalue of $\Delta_i F$ with eigenvector v. I will show that $\sigma(F) \geq |\lambda|$ by analyzing three distinct cases.

1. If $|\lambda| = 1$, then Lemma 1.7 implies that $\lim_{t\to\infty}\|\Delta_i F^t v\| = \lim_{t\to\infty}\|(\Delta_i F)^t v\| = \lim_{t\to\infty}\|\lambda^t v\| = \|v\| \neq 0$. Additionally, if $\sigma(F) < 1$, then Puterman (2005, p. 593, equation A6) and Lemma 1.1 imply that $\lim_{t\to\infty}\Delta_i F^t = 0$, which is a contradiction. Accordingly, $\sigma(F) \geq |\lambda|$ when $|\lambda| = 1$.

2. If $v \in \mathrm{span}(\iota_i)$, then Lemmas 1.1 and 1.3 imply that $\lambda v = \Delta_i F v = \Delta_i(\Delta_i F v) = \lambda\Delta_i v = 0$; this implies that $\lambda = 0$, which implies that $\sigma(F) \geq |\lambda|$.

3. If $|\lambda| \neq 1$ and $v \notin \mathrm{span}(\iota_i)$, then Lemmas 1.1 and Lemma 1.7 imply that λ is also an eigenvalue of F:

$$\begin{aligned}F(v-\alpha\iota_i) &= \Delta_i F(v-\alpha\iota_i) + (I_i - \Delta_i)F(v-\alpha\iota_i)\\ &= \Delta_i F\Delta_i(v-\alpha\iota_i) + (I_i-\Delta_i)Fv - \alpha(I_i-\Delta_i)F\iota_i\\ &= \Delta_i F\Delta_i v + (I_i-\Delta_i)Fv - \alpha(I_i-\Delta_i)\iota_i\\ &= \Delta_i F v + (I_i-\Delta_i)Fv - \alpha\iota_i\\ &= \lambda v + \iota_i\delta_i(1)'Fv - (1-\lambda)^{-1}(\delta_i(1)'Fv)\iota_i\\ &= \lambda v - \lambda(1-\lambda)^{-1}(\delta_i(1)'Fv)\iota_i\\ &= \lambda(v-\alpha\iota_i),\end{aligned}$$

where $\alpha = (1-\lambda)^{-1}\delta_i(1)'Fv$.

This implies that $\sigma(F) \geq |\lambda|$, because $\sigma(F)$ is as large as any eigenvalue of F with a modulus that does not equal one (Puterman 2005, p. 595). □

Lemma 7. *$\sigma(F) = \phi(\Omega F)$ for any $Y \times Y$ stochastic matrix F.*

Proof. Lemmas 1.6 and 1.9 imply that, if λ is an eigenvalue of $\Delta_Y F$ with eigenvector v, then λ is an eigenvalue of ΩF with eigenvector Ωv: $\Omega F \Omega v = \Omega F v = \Omega \Delta_Y F v = \lambda \Omega v$. This implies that $\phi(\Omega F) \geq \phi(\Delta_Y F)$. With this, Lemma 6 implies the result. □

Proof of Proposition 4. The proof has three parts.

1. Lemmas 5 and 6 establish that $\phi(\Lambda F(p)) \leq \sigma(F(p))$.
2. Lemmas 1.5 and 3 imply that $\widetilde{v}^t 0 = \check{v}\bar{v}^{t-1}0$, where $\widetilde{v} = \Delta v \Lambda$. Accordingly, $\bar{v}^t 0$ converges to $\bar{V}^*$ at the same rate as $\widetilde{v}^t 0$ converges to $\check{v}\bar{V}^*$. Note that $\widetilde{v}$ is equivalent to the relative value iteration operator under state transition matrix $\widetilde{F}(p) = (\alpha' \iota)^{-1}(\iota \alpha' - \Delta F(p)\Lambda)$ and discount factor $\widetilde{\beta} = \beta \alpha' \iota$, where α is the vector with the ith element that equals the minimum value in the ith column of $\Delta F(p)\Lambda$. Accordingly, the strong convergence result of Morton and Wecker (1977) and Bray (2019) implies that $\|\check{v}\bar{V}^* - \widetilde{v}^t 0\|$ is $O^+(\widetilde{\beta}^t \sigma(\widetilde{F}(p^*))^t)$ as $t \to \infty$. Lemmas 1.1, 1.3, 1.5, and 6 imply that

$$
\begin{aligned}
\widetilde{\beta}\sigma\left(\widetilde{F}(p^*)\right) &= \widetilde{\beta}\phi\left(\Delta \widetilde{F}(p^*)\right) \\
&= \beta \alpha' \iota \phi((\alpha' \iota)^{-1} \Delta(\iota \alpha' - \Delta F(p^*)\Lambda)) \\
&= \beta \phi\left(\Delta F(p^*)\Lambda\right) \\
&= \beta \phi\left(\Lambda \Delta F(p^*)\right) \\
&= \beta \phi\left(\Lambda F(p^*)\right).
\end{aligned}
$$

theorem 1.3.22 of Horn and Johnson (2013) establishes the equivalence of $\phi(\Delta F(p)\Lambda)$ and $\phi(\Lambda \Delta F(p))$.

3. Lemma 2 implies that $\bar{\eta}_t p = \Lambda \eta_t p$. With this, Lemmas 1.4 and 1.8 imply that $\bar{\eta} p - \bar{\eta}_t p = \Lambda(\eta p - \eta_t p) = \Lambda \sum_{\tau=t+1}^{\infty} \beta^\tau F(p)^\tau U(p) = \beta^{t+1} \Lambda F(p)^{t+1} \sum_{\tau=0}^{\infty} \beta^\tau F(p)^\tau U(p) = \beta^{t+1} \Lambda F(p)^{t+1} \eta p = \beta^{t+1}(\Lambda F(p))^{t+1} \eta p$, which is $O^+(\beta^t \phi \cdot (\Lambda F(p))^t)$ as $t \to \infty$. □

Proof of Proposition 5. Lemmas 4 and 6 imply that $\sigma(F(p)) \geq \sigma(F_x)$. Therefore, I just have to show that $\phi(\Lambda F(p)) \leq \max_{x \in \mathbb{x}} \|\Omega F_y(p, x)\|$. Let $|||\cdot|||$ be a vector norm that is compatible with matrix norm $\|\cdot\|$ (Horn and Johnson 2013, p. 347). Also, let λ be an eigenvalue of $\Lambda F(p)$ with corresponding eigenvector v. Note that $v = \sum_{i=1}^{X} \delta_X(i) \otimes v_i$, where $v_i = (\delta_X(i)' \otimes I_Y)v$. This implies that

$$
\begin{aligned}
|\lambda| |||v_i||| &= |||(\delta_X(i)' \otimes I_Y)\Lambda F(p) v||| \\
&= \Bigg|\Bigg|\Bigg| (\delta_X(i)' \otimes I_Y)(I_X \otimes \Omega) \\
&\quad \cdot \left(\sum_{k=1}^{X} (\delta_X(k)\delta_X(k)' F_x) \otimes F_y(p, x_k) \right) \\
&\quad \cdot \left(\sum_{j=1}^{X} \delta_X(j) \otimes v_j \right) \Bigg|\Bigg|\Bigg| \\
&= \Bigg|\Bigg|\Bigg| \sum_{j=1}^{X} \sum_{k=1}^{X} (\delta_X(i)' \delta_X(k) \delta_X(k)' F_x \delta_X(j)) \\
&\quad \otimes (\Omega F_y(p, x_k) v_j) \Bigg|\Bigg|\Bigg| \\
&= \Bigg|\Bigg|\Bigg| \sum_{j=1}^{X} (\delta_X(i)' F_x \delta_X(j)) \otimes (\Omega F_y(p, x_i) v_j) \Bigg|\Bigg|\Bigg| \\
&\leq \sum_{j=1}^{X} \delta_X(i)' F_x \delta_X(j) |||\Omega F_y(p, x_i) v_j||| \\
&\leq \sum_{j=1}^{X} \delta_X(i)' F_x \delta_X(j) \|\Omega F_y(p, x_i)\| \, |||v_j||| \\
&\leq \sum_{j=1}^{X} \delta_X(i)' F_x \delta_X(j) \, \|\Omega F_y(p, x_i)\| \max_k |||v_k||| \\
&= \|\Omega F_y(p, x_i)\| \max_k |||v_k|||, \\
&\leq \max_j \|\Omega F_y(p, x_j)\| \max_k |||v_k|||.
\end{aligned}
$$

This implies that $|\lambda| \leq \max_{x \in \mathbb{x}} \|\Omega F_y(p, x)\|$. □

Proof of Corollary 2. These conditions are simply Proposition 5 evaluated under different matrix norms. The first corresponds to the Frobenius norm, the second corresponds to the ℓ_2 norm, the third corresponds to the ℓ_∞ norm, the fourth corresponds to the ℓ_1 norm, and the fifth corresponds to the Hajnal matrix seminorm induced by the span vector seminorm (Puterman 2005, p. 198). The results hold under the Hajnal seminorm, because the eigenvector associated with the largest eigenvalue of $\Lambda F(p)$ has strictly positive span; this eigenvector has strictly positive span, because it is not a multiple of ι, and it is not a multiple of ι, because ι is an eigenvector of $\Lambda F(p)$ with zero eigenvalue. □

Proof of Corollary 3. If F_x is not ergodic, then $\sigma(F_x) = 1$ (Puterman 2005, p. 595). Also, if $F_y(p^*, x)$ and $F_y(p, x)$ are scrambling, then the expressions in (1e) and (2e) of Corollary 2 are less than one. □

Proof of Corollary 4. For any $\epsilon > 0$, there exists a matrix norm $\|\cdot\|_\epsilon$ such that $\|\Omega F_y(p)\|_\epsilon \leq \phi(\Omega F_y(p)) + \epsilon$ (Horn and Johnson 2013, p. 347). With this, Proposition 5 and Lemma 7 imply the result. □

Acknowledgments

The author thanks Arvind Magesan, Victor Aguirregabiria, and three anonymous reviewers.

Endnotes

[1] My approach extends to the more general specification in which $f(x', y'|a, x, y) = f_x(x'|x)f_y(y'|a, x, x', y)$. In this case, the state transition matrix defined in Section 3 changes to $F(p) = \sum_{i=1}^{X}\sum_{j=1}^{X}(\delta_X(i)\cdot \delta_X(i)' F_x(\delta_X(j)\delta_X(j)') \otimes F_y(p, x_i, x_j)$, where F_x is the $X \times X$ matrix with ijth element $f_x(x_j|x_i)$ and $F_y(p, x, x')$ is the $Y \times Y$ matrix with ijth element $f_y(y_j|p(x, y_i), x, x', y_i)$.

[2] I assume that πV is unique.

[3] Bray (2019) actually provided the first valid proof of the first result.

[4] Note that $\mathbb{V}_x$ does not store all of the information about x; it stores only the information about x that is independent of y.

[5] Note that ergodicity in $F_y(p, x)$, for each $x \in \mathbb{x}$, does not imply ergodicity in y. For example, if

$$F_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad F_y(p, x_1) = \begin{bmatrix} 0 & 1 \\ .5 & .5 \end{bmatrix}, \quad \text{and}$$
$$F_y(p, x_2) = \begin{bmatrix} .5 & .5 \\ 1 & 0 \end{bmatrix},$$

then y is periodic, although $F_y(p, x_1)$ and $F_y(p, x_2)$ are both ergodic. However, y must be ergodic if $F_y(p, x)$ is scrambling for each $x \in \mathbb{x}$ (Seneta 2006, p. 80).

[6] Paarsch and Rust (2009, p. 1) explain that many dynamic programs in economics contain such a seasonal component: "Many dynamic programming problems encountered in practice involve a mix of state variables, some exhibiting stochastic cycles (such as unemployment rates) and others having deterministic cycles. Examples of the latter include the day of the week as well as the month and the season of the year.. . .Most real-world problems involve complicated interactions between variables that evolve according to deterministic cycles and those that evolve according to stochastic cycles. In many nonlinear models, no simple method exists to isolate the deterministically evolving components from the stochastically evolving ones, especially when agents are responding endogenously to both kinds of components."

[7] Following Tauchen (1986), I set each factor's five values to the sextiles of the underlying AR(1) process's unconditional distribution.

[8] Chen (2017) reported similar results, explaining that endogenous value iteration solved her acid rain dynamic programs 5.2 times faster than relative value iteration and 83 times faster than standard value iteration.

References

Adelman D, Mancini AJ (2016) Optimality of quasi-open-loop policies for discounted semi-Markov decision processes. *Math. Oper. Res.* 41(4):1222–1247.

Aguirregabiria V, Magesan A (2013) Euler equations for the estimation of dynamic discrete choice structural models. *Adv. Econometrics* 31(46056):1–42.

Aguirregabiria V, Magesan A (2018) Solution and estimation of dynamic discrete choice structural models using Euler equations. Working paper, University of Toronto, Toronto.

Aguirregabiria V, Mira P (2010) Dynamic discrete choice structural models: A survey. *J. Econometrics* 156(1):38–67.

Bray RL (2019) Strong convergence and dynamic economic models. *Quant. Econom.* Forthcoming.

Chen C (2017) Slow focus: Belief evolution in the U.S. acid rain program. Working paper, John F. Kennedy School of Government, Cambridge, MA.

Hajnal J (1957) Weak ergodicity in non-homogeneous Markov chains. *Math. Proc. Cambridge Philos. Soc.* 54(2):233–246.

Higle JL, Bean JC, Simth RL (1990) Deterministic equivalence in stochastic infinite horizon problems. *Math. Oper. Res.* 15(3): 396–407.

Horn RA, Johnson CR (2013) *Matrix Analysis* (Cambridge University Press, New York).

Morton TE (1971) On the asymptotic convergence rate of cost differences for Markovian decision processes. *Oper. Res.* 19(1): 244–248.

Morton TE, Wecker WE (1977) Discounting, ergodicity and convergence for Markov decision processes. *Management Sci.* 23(8): 890–900.

Paarsch HJ, Rust J (2009) Valuing programs with deterministic and stochastic cycles. *J. Econom. Dynam. Control* 33(3): 614–623.

Puterman ML (2005) *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (John Wiley & Sons, Hoboken, NJ).

Seneta E (2006) *Non-Negative Matrices and Markov Chains* (Springer Science & Business Media, New York).

Tauchen G (1986) Finite state Markov-chain approximations to univariate and vector autoregressions. *Econom. Lett.* 20(2): 177–181.