Convergence of Discrete-Time Models with Small Period Lengths

Joel Watson*

April 2021

Abstract

This paper provides uniform convergence results that relate (i) a sequence in discrete time that is defined inductively with respect to a transition function and (ii) the solution of a differential initial-value problem in continuous time that is derived from the same transition function. The results are useful for characterizing dynamics of some economic models.

1 Introduction

Consider a discrete-time modeling exercise with a solution that can be expressed inductively as a sequence of states on the real line. The model specifies an initial state $\underline{x} \in \mathbb{R}$, and a period length (step size) $\Delta > 0$. The solution (the model's prediction) is described by a sequence $\{x^k(\Delta)\}_{k=0}^{\infty}$, where x^k denotes the state in period k. Suppose that the solution is characterized by a function $f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, such that

$$x^0(\Delta) = \underline{x} \quad \text{and} \quad x^{k+1}(\Delta) = f(x^k(\Delta); \Delta) \quad \text{for each } k \in P,$$

where $P = \{0, 1, 2, \ldots\}$ is the set of whole numbers.[1]

In some settings, it would be inconvenient or difficult to solve for $\{x^k(\Delta)\}$. Nevertheless, it may be possible to calculate a continuous-time limit as $\Delta \to 0^+$, by using the transition function to characterize the slope $(x^{k+1} - x^k)/\Delta$. Specifically, suppose that

$$g(x) \equiv \lim_{\Delta \to 0^+} \frac{f(x; \Delta) - x}{\Delta} \tag{1}$$

exists for every x. Note that if f is continuously differentiable, then from L'Hôpital's rule the limit defining g exists if and only if $f(x, 0) = x$, and we have $g(x) = \frac{\partial f}{\partial \Delta}(x; 0)$.

For a given time horizon $T > 0$, we may look for a solution $y\colon [0, T] \to \mathbb{R}$ to the differential equation $y' = g(y)$ with initial value $y(0) = \underline{x}$. Here y' denotes the first derivative of y. Assuming there is a unique solution to the initial-value problem, a key issue issue is whether y is a good approximation of $\{x^k(\Delta)\}$ for small values of Δ. To

*UC San Diego, https://joelwatson.org. Watson thanks Xiamen Hua, Freddie Papazyan, Joel Sobel, Maxwell Stinchcombe, and Endre Süli for suggestions and Haitian Xie for research assistance.

[1] Time-based terminology (such as "period length" rather than "step size") is used here because it suits many of the economic applications.

make this precise, let us convert the sequence into a function that gives the state x as a function of time t on the continuum, so that it is comparable to y. For simplicity, let it be the step function defined by

$$\hat{x}(t;\Delta) \equiv x^{[t/\Delta]}(\Delta),$$

where for any number $w \geq 0$, $[w]$ denotes the largest integer k satisfying $k \leq w$. We are interested in knowing whether $\hat{x}(\cdot;\Delta)$ converges to $y(t)$ *uniformly in* $t \in [0,T]$ as Δ approaches zero. If so, we can say that the function y well approximates the sequence $\{x^k(\Delta)\}_{k=0}^{[T/\Delta]}$ on $[0,T]$.

This paper examines a general version of the setting just described and establishes that convergence is assured under appropriate conditions on the primitives that one would hope to be able to check in applications. In the general setting, x^{k+1} is a function of both x^k and a number of lagged states $x^{k-1}, x^{k-2}, \ldots, x^{k-L}$ for some integer L. The first $L+1$ elements of the sequence are given and can depend on Δ, so long as they converge to $\underline{x}$ as $\Delta \to 0^+$. Thus the solution sequence is specified by:

$$\begin{aligned} &x^{\ell}(\Delta) \text{ given, for } \ell = 0, 1, \ldots, L, \text{ and} \\ &x^{k+1}(\Delta) = f(x^k(\Delta), x^{k-1}(\Delta), \ldots, x^{k-L}(\Delta); \Delta) \text{ for each } k = L, L+1, \ldots. \end{aligned} \tag{2}$$

The transition function is assumed to be twice continuously differentiable and to satisfy $f(x, x, \ldots, x; 0) = x$ for every x.

Let f_0 denote the derivative of f with respect to its first argument, x^k, let f_1 denote the derivative of f with respect to its second argument, x^{k-1}, and so on up to f_L denoting the derivative with respect to the $L+1$ argument, x^{k-L}. Let f_Δ denote the derivative of f with respect to its last argument, Δ. The main theorem presented here establishes that, with appropriate bounds on fundamentals for states in a suitable subset of $\mathbb{R}$, $\hat{x}(\cdot;\Delta)$ converges uniformly to the function $y\colon [0,T] \to \mathbb{R}$ that solves the differential equation $y' = g(y)$ with initial value $y(0) = \underline{x}$, where g is defined by

$$g(x) \equiv \frac{f_\Delta(x, x, \ldots, x; 0)}{1 + \sum_{\ell=1}^{L} \ell\, f_\ell(x, x, \ldots, x; 0)}, \tag{3}$$

for every x. Further, there is a unique solution to the initial-value problem.

The analysis presented here is closely related to the literature on numerical methods to approximate the solution of an intractable initial-value problem, where the starting point is a given function g and the goal is to solve the differential equation $y' = g(y)$. Euler's one-step method for approximating the solution is to partition the time interval $[0,T]$ and then construct a spline function $\hat{z}$ whose slope in each sub-interval is given by g evaluated at the sub-interval's left bound. For subintervals of equal size Δ, Euler's method essentially constructs the sequence $\{z^k(\Delta)\}$ defined by $z^0 = \underline{x}$ and $z^{k+1}(\Delta) = z^k(\Delta) + g(z^k(\Delta))\Delta$.

In comparison, this paper examines the reverse direction, where the starting point is a function f that characterizes the solution of a discrete-time model, and the goal is to represent its limit as the period length shrinks. Conceptually, it is useful to think of both $\hat{x}$ and $\hat{z}$ as candidate approximations of y. Standard results in the numerical-methods literature establish the relation between $\hat{z}$ and y. The contribution here is to provide the form of g for a given f and establish convergence of $\hat{x}$ to y, which in the proof is actually accomplished by relating $\hat{x}$ to $\hat{z}$. The analysis is complicated by the fact that $x^k(\Delta)$ is not guaranteed to be close to $z^k(\Delta)$ for possibly many values of k, due to the arbitrary nature of the initial values $x^0(\Delta), x^1(\Delta), \ldots, x^L(\Delta)$. But it turns out that the approximation becomes incrementally better for large k, and this offsets the poor approximation for smaller values of k as $\Delta \to 0^+$.

Also reported here is a second theorem that allows the state to be a vector in a Euclidean space $\mathbb{R}^n$ but assumes $L = 0$, so no lagged states are arguments in the transition function. The functions f, g, and y are now vector-valued. Dispensing with lagged states makes the extension is easier to describe, and the proof greatly simplified, in comparison to the main theorem, but it also narrows applicability.

It is natural question to ask whether the second theorem actually subsumes the main theorem, because we know that any dynamical system with lagged variables can be transformed into a multidimensional system without lags. The answer is *negative*, however, because assumptions needed for convergence in the former setting do not imply the assumptions needed in the latter setting. The main theorem and its extension therefore serve different applications. Details are provided following the statement of the results in the next section.

The results may be useful for applications in four ways. First, they provide a convenient characterization of the solution to a discrete-time model when the differential equation is easier to solve. This can be the case if the function f is difficult to deal with for positive values of Δ, but derivatives can be calculated at $\Delta = 0$. Another case is where one can deal with the function f, yet the initial values $x^0, x^1, \ldots, x^L$ depend on Δ and are not easily characterized.

Second, the discrete-time model may have multiple solutions, identified by various combinations of initial values and/or transition functions. In such a case, one may be able to bound the solutions using a particular transition function and specification of initial values. Third, even if the discrete-time solution can be computed without much difficulty, it may be helpful to know that it convergences uniformly as the period length shrinks.

Finally, the results provide a new option for numerical methods. If the objective is to characterize a discrete-time dynamical system defined by $f(\cdot;\Delta)$ for a small period length, the results establish that performing Euler's method using the related function f is a good approximation.

The main theorem is presented formally in the next section, which begins with the development of a little intuition. The second theorem is presented at the end of the section, along with a discussion of the relation between the two results. The section after describes calculations in four economic examples: a standard growth model, a model of knowledge acquisition by research and development, a model of power dynamics, and a recurrent contracting setting with privately information. The final section contains a proof of the theorem, with six lemmas at the core. There is no concluding section because it seems unnecessary to include one.

2 Theorems

Before stating the main theorem, it is useful to present some intuition. Consider the case of $L = 2$ and take a point in continuous time $t \in \mathbb{R}_+$. For any Δ, the period closest to t is $k = [t/\Delta]$. Let us describe the slope

$$\frac{\hat{x}(t+\Delta;\Delta) - \hat{x}(t;\Delta)}{\Delta} = \frac{x^{k+1}(\Delta) - x^k(\Delta)}{\Delta}$$

and give a heuristic characterization of the limit as Δ converges to 0. Using Equation 2, this slope is

$$\frac{f(x^k(\Delta), x^{k-1}(\Delta), x^{k-2}(\Delta); \Delta) - x^k(\Delta)}{\Delta}.$$

For convenience, imagine that $x^k(\Delta)$ is a constant x. The key thing to notice is that $x^{k-1}(\Delta)$ and $x^{k-2}(\Delta)$ change as Δ approaches 0, and for the discrete-time model to converge, it must be that $x^{k-1}(\Delta)$ and $x^{k-2}(\Delta)$ converge to x.

Then as Δ approaches 0, its direct effect on the slope is roughly

$$\lim_{\Delta \to 0^+} \frac{f(x,x,x;\Delta) - x}{\Delta} = f_\Delta(x,x,x;0).$$

Its effect through $x^{k-1}(\Delta)$ is approximately

$$\lim_{\Delta \to 0^+} \frac{f(x, x^{k-1}(\Delta), x; 0) - x}{\Delta} = -f_1(x,x,x;0) \cdot \lim_{\Delta \to 0^+} \frac{x - x^{k-1}(\Delta)}{\Delta},$$

and its effect through $x^{k-2}(\Delta)$ is roughly

$$\lim_{\Delta \to 0^+} \frac{f(x, x, x^{k-2}(\Delta); 0) - x}{\Delta} = -f_2(x,x,x;0) \cdot \lim_{\Delta \to 0^+} \frac{x - x^{k-2}(\Delta)}{\Delta}.$$

Writing $x - x^{k-2}(\Delta) = \left(x - x^{k-1}(\Delta)\right) + \left(x^{k-1}(\Delta) - x^{k-2}(\Delta)\right)$, we thus conclude that the slope of y at time t is

$$\begin{aligned} g(x) \equiv f_\Delta(x,x,x;0) - f_1(x,x,x;0) \cdot \lim_{\Delta \to 0^+} \frac{x - x^{k-1}(\Delta)}{\Delta} \\ - f_2(x,x,x;0) \cdot \lim_{\Delta \to 0^+} \left(\frac{x - x^{k-1}(\Delta)}{\Delta} + \frac{x^{k-1}(\Delta) - x^{k-2}(\Delta)}{\Delta} \right). \end{aligned}$$

Further, $\lim_{\Delta \to 0^+} \left(x - x^{k-1}(\Delta)\right)/\Delta$ and $\lim_{\Delta \to 0^+} \left(x - x^{k-1}(\Delta)\right)/\Delta$ must be the same slope $g(x)$. So we have

$$g(x) = f_\Delta(x,x,x;0) - f_1(x,x,x;0)g(x) - 2f_2(x,x,x;0)g(x).$$

Solving for $g(x)$ yields Equation 3 for the case of $L = 2$.

Observe that in the case of $L > 1$, the ℓth lagged state has weight ℓ in the summation of derivatives that defines g. For intuition, note that it is the differences between states in prior periods and the state in the current period that affects the magnitude by which the state changes from the current period to the next. For example, suppose that in the solution, the state gradually increases over successive periods of time. Then the difference between x^k and x^{k-2} will be roughly twice the difference between x^k and x^{k-1}; hence the effect of f_2 will be roughly twice the effect of f_1 in determining the state in period $k+1$, and similarly for larger values of ℓ.

Theorem 1: *Take the following as given to define, for every $\Delta \in \mathbb{R}$, a sequence of real numbers denoted $\{x^k(\Delta)\}_{k=0}^{\infty}$:*

- *number of lagged states $L \in P$;*
- *transition function $f\colon \mathbb{R}^{L+1} \times \mathbb{R} \to \mathbb{R}$;*
- *initial-state functions $x^\ell\colon \mathbb{R} \to \mathbb{R}$ for $\ell = 0, 1, \ldots, L$;*
- *limit initial state $\underline{x} \in \mathbb{R}$; and*
- *time horizon $T > 0$.*

Given Δ, the sequence $\{x^k(\Delta)\}_{k=0}^{\infty}$ is defined inductively by

$$x^{k+1}(\Delta) = f(x^k(\Delta), x^{k-1}(\Delta), \ldots, x^{k-L}(\Delta); \Delta),$$

for every $k \geq L$.

Assumptions: Let X be an open interval of $\mathbb{R}$ containing $\underline{x}$. Let Γ be an open interval of $\mathbb{R}$ containing 0. Assume f is twice continuously differentiable on $X^{L+1} \times \Gamma$, $f(x, x, \ldots, x; 0) = x$ for every $x \in X$, and there is a number $a < 1$ such that $\sum_{\ell=1}^{L} \ell\, |f_\ell(x, x, \ldots, x; 0)| \leq a$ for every $x \in X$. Define $g\colon X \to \mathbb{R}$ by

$$g(x) \equiv \frac{f_\Delta(x, x, \ldots, x; 0)}{1 + \sum_{\ell=1}^{L} \ell\, f_\ell(x, x, \ldots, x; 0)}.$$

Suppose there is a number $A > 0$ such that $|g(x)| \leq A$ for all $x \in [\underline{x} - TA, \underline{x} + TA] \subset X$. Assume also that $\big(x^\ell(\Delta) - x^{\ell-1}(\Delta)\big)/\Delta$ is bounded for every $\ell = 1, 2, \ldots, L$, and that $x^0(\Delta)$ converges to $\underline{x}$ as $\Delta \to 0^+$.

Conclusion: Define $\hat{x}\colon [0, T] \times \mathbb{R}_{++} \to \mathbb{R}$ by $\hat{x}(t; \Delta) \equiv x^{[t/\Delta]}(\Delta)$. Then there is a unique solution $y\colon [0, T] \to X$ to the initial-value problem $y' = g(y)$ with $y(0) = \underline{x}$, and $\hat{x}(\,\cdot\,; \Delta)$ converges uniformly to y as $\Delta \to 0^+$.

The next result extends the main theorem in the case of $L = 0$ by allowing the state to be a vector in $\mathbb{R}^n$. For vector-valued functions, let superscripts refer to the component of the output vector, hopefully not to be confused with the index of sequences. Continue to write subscript Δ to denote the derivative with respect to Δ and let prime denote the first derivative of a function of the reals. For differentiable functions $f\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ and $y\colon [0, T] \to \mathbb{R}^n$, we can therefore write

$$y(t) = \begin{pmatrix} y^1(t) \\ y^2(t) \\ \vdots \\ y^n(t) \end{pmatrix}, \quad y' = \begin{pmatrix} (y^1)' \\ (y^2)' \\ \vdots \\ (y^n)' \end{pmatrix}, \quad \text{and } f_\Delta(x; \Delta) = \begin{pmatrix} \frac{\partial f^1}{\partial \Delta}(x; \Delta) \\ \frac{\partial f^2}{\partial \Delta}(x; \Delta) \\ \vdots \\ \frac{\partial f^n}{\partial \Delta}(x; \Delta) \end{pmatrix}.$$

Let $\|\cdot\|$ denote the Euclidean norm.

Theorem 2: *Let n be a positive integer. Take the following as given to define, for every $\Delta \in \mathbb{R}$, a sequence of vectors in $\mathbb{R}^n$ denoted $\{x^k(\Delta)\}_{k=0}^{\infty}$:*

- *transition function $f\colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$;*
- *initial-state function $x^0\colon \mathbb{R} \to \mathbb{R}^n$;*
- *limit initial state $\underline{x} \in \mathbb{R}^n$; and*
- *time horizon $T > 0$.*

Given Δ, the sequence $\{x^k(\Delta)\}_{k=0}^{\infty}$ is defined inductively by $x^{k+1}(\Delta) = f(x^k(\Delta); \Delta)$ for every $k \geq L$.

Assumptions: Let $X \subset \mathbb{R}^n$ be an open ball containing $\underline{x}$. Let Γ be an open interval of $\mathbb{R}$ containing 0. Assume f is twice continuously differentiable on $X \times \Gamma$, assume $f(x; 0) = x$ for every $x \in X$, and define $g\colon X \to \mathbb{R}^n$ by $g(x) \equiv f_\Delta(x; 0)$. Suppose there is a number $A > 0$ such that $\|x - \underline{x}\| \leq TA$ implies $x \in X$ and $\|g(x)\| \leq A$. Assume also that $x^0(\Delta)$ converges to $\underline{x}$ as $\Delta \to 0^+$.

Conclusion: Define $\hat{x}\colon [0, T] \times \mathbb{R}_{++} \to \mathbb{R}^n$ by $\hat{x}(t; \Delta) \equiv x^{[t/\Delta]}(\Delta)$. Then there is a unique solution $y\colon [0, T] \to X$ to the initial-value problem $y' = g(y)$ with $y(0) = \underline{x}$, and $\hat{x}(\,\cdot\,; \Delta)$ converges uniformly to y as $\Delta \to 0^+$.

One may be inclined to think that Theorem 2 subsumes Theorem 1, because any dynamical system with lagged variables can be transformed into a multidimensional system without lags, but this is not the case. Consider, for example, a setting with $L = 1$ where $x^{k+1}(\Delta) = f(x^k(\Delta), x^{k-1}(\Delta); \Delta)$. Suppose the assumptions of Theorem 1 are satisfied. To transform this system into one with no lags, we define a sequence in $\mathbb{R}^2$, denoted $\{\tilde{x}^k(\Delta)\}$, by specifying

$$\tilde{x}^k(\Delta) = \begin{pmatrix} \tilde{x}_1^k(\Delta) \\ \tilde{x}_2^k(\Delta) \end{pmatrix} \equiv \begin{pmatrix} x^k(\Delta) \\ x^{k-1}(\Delta) \end{pmatrix}$$

and letting the transition function be

$$\tilde{f}(\tilde{x}^k; \Delta) \equiv \begin{pmatrix} f(\tilde{x}_1^k(\Delta), \tilde{x}_2^k(\Delta); \Delta) \\ \tilde{x}_1^k(\Delta) \end{pmatrix}.$$

Note, however, that $\tilde{f}(\tilde{x}; \Delta)$ does not equal $\tilde{x}$ generally, so a key assumption of Theorem 2 do not necessarily hold, and the derivatives $\tilde{f}$ are insufficient for characterizing the sequence as Δ becomes small. Evaluating the slope in the second vector component, $(\tilde{x}_2^{k+1}(\Delta) - \tilde{x}_2^k(\Delta))/\Delta$, is the same as evaluating $(x^k(\Delta) - x^{k-1}(\Delta))/\Delta$. To provide an estimate of $(x^{k+1}(\Delta) - x^k(\Delta))/\Delta$, not only must we calculate how $f(x^k, x^{k-1}; \Delta)$ varies with Δ but we also must incorporate how close x^{k-1} is to x^k and thus how $f(x^k, x^{k-1}; \Delta)$ varies with x^{k-1}. This shows why settings with $L > 0$ involve challenging subtleties and, further, that Theorems 1 and 2 serve different applications.

3 Examples

Example 1: growth model

Let us start with a trivial example, the neoclassical economic growth model of Solow (1956) and Swan (1956). The basic model in discrete time yields the following equation governing the sequence of capital intensity $\{x^k\}$, where parameter s is the savings rate, λ is the capital depreciation rate, and function h represents equilibrium utilization of the production technology:

$$x^{k+1} = sh(x^k)\Delta + (1 - \Delta\lambda)x^k \equiv f(x^k; \Delta).$$

Note that in this example the state is a real number and $L = 0$. We have $f(x; 0) = x$ and $g(x) = f_\Delta(x; 0) = sh(x^k) + -\lambda x^k$.

In the continuous-time version of the model, the state evolves over time according to the differential equation $y' = f_\Delta(y; 0) = sh(y) - \lambda y$. Theorem 1 applies for a given initial capital intensity and suitable time interval, assuming h is well behaved. The path of capital intensity over time in the discrete model converges uniformly to the path identified in the continuous-time model. Thus, the two versions of the model are consistent not only in terms of the steady states but also in precisely the rate that capital intensity changes over time from any initial condition.

Example 2: model of knowledge acquisition

Consider a simple setting in which an organization engages in research and development to build a stock of knowledge capital. Time is in discrete periods $k = 0, 1, \ldots$ with period length Δ. The stock of knowledge at the end of period k is denoted by x^k. The initial stock, at the end of periods 0 and 1, is $x^0 = x^1 \equiv 1$. Activity begins in period 2. In each period $k + 1$, managers of the organization allocate any amount of their existing capital to fund projects that will increase the stock.

Two projects are available, an *existing project* whose returns depend on the recent rate that knowledge has been increasing, $(x^k - x^{k-1})/\Delta$, and a *new project* whose returns depend on the stock x^k at the beginning of the period. An amount w_E allocated to the existing project returns capital in the amount of

$$2(w_\mathrm{E}\Delta\beta)^{\frac{1}{2}} \cdot \left(\frac{x^k - x^{k-1}}{\Delta}\right)^{\frac{1}{2}},$$

whereas w_N allocated to the new project generates

$$2(w_\mathrm{N}\Delta\alpha x^k)^{\frac{1}{2}},$$

for a total gain of

$$h(w_\mathrm{E}, w_\mathrm{N}, x^k, x^{k-1}; \Delta) = 2(w_\mathrm{E}\Delta\beta)^{\frac{1}{2}} \cdot \left(\frac{x^k - x^{k-1}}{\Delta}\right)^{\frac{1}{2}} + 2(w_\mathrm{N}\Delta\alpha x^k)^{\frac{1}{2}} - w_\mathrm{E} - w_\mathrm{N}.$$

The superscript on x refers to the period, whereas the other superscripts are exponents, and parameters α and β are positive numbers. Decreasing returns are implied by w_{E} and w_{N} having the exponent 1/2. The inclusion of $\Delta^{1/2}$ in these production terms is needed to appropriately scale by period length, so that optimal choices in a period of length Δ imply a capital gain that is Δ times the gain for a period of length one.

Our objective is to characterize the capital stock over time under the assumption that management maximizes growth in each period. The first-order conditions for maximization yield optimal allocations $w^*_{\mathrm{E}} = \beta(x^k - x^{k-1})$ and $w^*_{\mathrm{N}} = \alpha\Delta x^k$, and we obtain $h(w^*_{\mathrm{E}}, w^*_{\mathrm{N}}, x^k, x^{k-1}; \Delta) = \beta(x^k - x^{k-1}) + \alpha\Delta x^k$. We have $L = 1$ and the capital stock transition is given by $x^{k+1} = f(x^k, x^{k-1}; \Delta)$ where

$$f(x^k, x^{k-1}; \Delta) \equiv x^k + \beta(x^k - x^{k-1}) + \alpha\Delta x^k.$$

Note that $f_\Delta(x, x; 0) = \alpha x$ and $f_1(x, x; 0) = -\beta$.

The set X can be taken to be any open, bounded interval that contains 1. For a suitably short time horizon T, Theorem 1 applies assuming $\beta < 1$. Then $\hat{x}(\cdot; \Delta)$ is uniformly approximated by $y\colon [0, T] \to \mathbb{R}$ solving $y' = g(y) = \alpha y/(1 - \beta)$ with initial value $y(0) = 1$. Integrating yields $y(t) = e^{\alpha t/(1-\beta)}$. Relative dynamics of new and existing project investments are easily derived. In the case of $\beta > 1$, returns of the existing project imply that the capital stock increases on order greater than Δ, implying the model is not well behaved as the period length becomes small, and the continuous-time model is not well defined.

Example 3: model of power dynamics

Here is an example of power dynamics between generations of competing political actors, similar to the model of Acemoglu and Robinson (2018). The model is discrete-time noncooperative game in which, in each of an infinite number of periods, players 1 and 2 choose how much to invest into their respective stocks of power. The power of player $i \in \{1, 2\}$ in period k is denoted x_i^k and its transition to period $k + 1$ is given by $x_i^{k+1} = x_i^k(1 - \Delta\lambda) + q_i^{k+1}$, where q_i^{k+1} is player i's investment in period $k + 1$ and λ is a parameter measuring the natural depletion rate of power. Player i's payoff in period $k + 1$ is

$$u_i^{k+1} = \alpha x_i^{k+1} + \beta x_j^{k+1} + \gamma x_i^{k+1} x_j^{k+1} - \frac{(q_i^{k+1})^2}{2\Delta},$$

where the last term is player i's investment cost (normalized by the period length) and the first three terms capture player i's benefit of power in relation to the power of the other player, denoted x_j^{k+1}. Numbers α, β, and γ are parameters.

Let us assume that the players in period $k + 1$ care only about the benefits and costs in this period, so player 1 chooses q_1^{k+1} to maximize u_1^{k+1} and player 2 chooses q_2^{k+1} to maximize u_2^{k+1}. Each player's optimal investment is a function of the other player's investment and the given levels x_1^k and x_2^k from the previous period. Solving the system of these two best-response functions identifies the unique Nash equilibrium for interaction in period k, which yields the following expression for the transition of power:

$$x^{k+1} = \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \end{pmatrix} = f(x^k; \Delta)$$
$$= \begin{pmatrix} \left[1 - \lambda\Delta + \frac{\gamma^2\Delta^2(1-\Delta\lambda)}{1-\gamma^2\Delta^2}\right] x_1^k + \left[\frac{\gamma\Delta(1-\Delta\lambda)}{1-\gamma^2\Delta^2}\right] x_2^k + \left[\frac{\alpha\Delta+\alpha\gamma\Delta^2}{1-\gamma^2\Delta^2}\right] \\ \left[1 - \lambda\Delta + \frac{\gamma^2\Delta^2(1-\Delta\lambda)}{1-\gamma^2\Delta^2}\right] x_2^k + \left[\frac{\gamma\Delta(1-\Delta\lambda)}{1-\gamma^2\Delta^2}\right] x_1^k + \left[\frac{\alpha\Delta+\alpha\gamma\Delta^2}{1-\gamma^2\Delta^2}\right] \end{pmatrix}.$$

Theorem 2 applies for this example, as we have a setting in which the state is in $\mathbb{R}^2$ and there are no lagged states in the transition function. Note that

$$f_\Delta(x; 0) = \begin{pmatrix} -\lambda x_1 + \gamma x_2 + \alpha \\ \gamma x_1 - \lambda x_2 + \alpha \end{pmatrix},$$

and so the related system of differential equations is given by

$$y' = \begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} -\lambda & \gamma \\ \gamma & -\lambda \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} \equiv My + Q.$$

Using the matrix-exponential method, any solution can be written in the form

$$y(t) = e^{Mt}\underline{x} + e^{Mt}\int_0^t e^{-rM}Qdr,$$

where $\underline{x}$ is the vector of initial power levels. Diagonalization of M allows us to write the solution directly by calculating the exponential of a diagonal matrix. This works if $\gamma \neq -\lambda$. Taking the case of $\gamma > 0$ for illustration, the eigenvalues of M are $-\lambda - \gamma$ and $\gamma - \lambda$, and $\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ is a matrix of eigenvectors. After a few mathematical steps, we obtain:

$$y_i(t) = \tfrac{1}{2}\left(e^{-(\lambda+\gamma)t} + e^{(\gamma-\lambda)t}\right)\underline{x}_i + \tfrac{1}{2}\left(e^{(\gamma-\lambda)t} - e^{-(\lambda+\gamma)t}\right)\underline{x}_j + \tfrac{\alpha}{\gamma-\lambda}e^{(\gamma-\lambda)t}\left(e^{(\gamma-\lambda)t} - 1\right),$$

for $i \in \{1, 2\}$. The discrete-time dynamics converge uniformly to the continuous-time version.

Example 4: model of starting small in a relationship

The discrete-time, incomplete-information model of Hua and Watson (2020), which continues the line of research on "starting small" initiated by Watson (1999, 2002), utilizes Theorem 1 to characterize upper and lower bounds on the set of perfect Bayesian equilibria (PBE). The model features a principal and agent who interact over an infinite number of discrete periods. The period length is denoted Δ as usual. Before the first period, nature chooses the agent's type $w \in \mathbb{R}_+$ according to a distribution that puts some probability on 0 (the "good" type) and the remaining probability uniformly on an interval $[\underline{w}, \bar{w}]$ (the "bad" types), where $\underline{w} > 1$. The agent privately observes her type. In each period k, the principal chooses a level of trust $\alpha \in [0,1]$ and the agent, observing α, then decides whether to cooperate or betray. Cooperation leads to the payoff of $\alpha\Delta$ for both players. Betrayal yields $-c\alpha$ to the principal and αw to the agent, where $c > 0$, and effectively the game ends. The interest rate is 1, implying a discount factor of $e^{-\Delta}$.

For Δ close to zero, cooperation at the highest level (or at any positive level) could be sustained in a complete-information environment if and only if $w = 0$. This is because, at any constant level α, betraying to get αw is more attractive for the agent than is cooperating forever, which pays $\alpha\Delta/(1 - e^{-\Delta})$, if and only if $w < 1$. It follows that in any "cooperative PBE," in which the good type cooperates forever, all bad types betray in bounded time. Further, higher types betray earlier than do lower types. Under an additional condition relating to renegotiation by the principal, every cooperative PBE can be characterized by an integer $K \in P$, an equilibrium sequence of levels $\{\alpha^k\}_{k=1}^{\infty}$, and the equilibrium sequence of "cutoff types" $\{w^k\}_{k=1}^{\infty}$, with the properties described next.

The number α^k is the level that the principal chooses in period k on the equilibrium path, conditional on the agent cooperating through period $k-1$. Sequence $\{\alpha^k\}$ is strictly increasing through period $K+1$, and then $\alpha^{K+1} = \alpha^{K+2} = \alpha^{K+3} = \cdots = 1$. Sequence $\{w^k\}$ is strictly decreasing until period K, when it falls below 1 and then remains constant. In the PBE, for each $k \in P$, types in the interval $(w^{k-1}, w^k]$ betray in period k. Further, in any period $k < K$, type w^k is indifferent between betraying in period k and waiting to betray in period $k+1$, which is captured by the following equation:

$$\alpha^k w^k = \alpha^k \Delta + e^{-\Delta}\alpha^{k+1} w^k. \tag{4}$$

The renegotiation condition also jointly constrains the rates at which $\{\alpha^k\}$ increases, $\{w^k\}$ decreases, and player 1's continuation value changes over time.

Hua and Watson (2020) characterize the set of cooperative PBE that satisfy their renegotiation condition, in the limit as the period length Δ approaches zero. The analysis is quite complicated and so the mathematical expressions are not presented here, but it is useful to note how Theorem 1 is utilized. The authors are able to combine conditions such as Equation 4, equilibrium identities, and the renegotiation concept to derive bounds on $\{w^k\}$, specifically (i) an upper bound on w^{k-1} as a function of x^k and Δ and (ii) a lower bound on w^{k-1} as a function of x^k, x^{k+1}, and Δ. The first defines a sequence with transition $w^{k-1} = \bar{f}(w^k; \Delta)$ and the second defines a sequence with transition $w^{k-1} = \underline{f}(w^k, w^{k+1}; \Delta)$. Letting $\{x^k\}_{k=0}^{K}$ be given by $x^k = w^{K-k}$, the bounding sequences are put in the form described here. Although the transition functions are complicated and require implicit differentiation, the authors show that Theorem 1 applies for both $\bar{f}$ and $\underline{f}$, and interestingly these functions have the same limit differential equation. Therefore, the equilibrium is unique in the limit.

4 Proofs

4.1 Proof the Main Theorem

This section contains a proof of Theorem 1. It is somewhat involved and is developed through a series of six lemmas.

Define $\Gamma_{++} \equiv \Gamma \cap \mathbb{R}_{++}$. Let the collection of sequences $\{x^k(\Delta)\}$ for $\Delta \in \Gamma_{++}$ be derived from function f, with initial values $x^0(\Delta), x^1(\Delta), \ldots, x^L(\Delta)$, as described in the Theorem, and let $\hat{x} \times \Gamma_{++} \colon [0, T] \to X$ be defined by $\hat{x}(t; \Delta) \equiv x^{[t/\Delta]}(\Delta)$. Let P_K^M denote the subset of integers $\{K, K+1, K+2, \ldots, M\}$ and let $P_K \equiv \{K, K+1, K+2, \ldots\}$.

Because X is open, there exists a number $\xi > TA$ such that $Y \equiv [\underline{x} - \xi, \underline{x} + \xi] \subset X$, and so we have $[\underline{x} - TA, \underline{x} + TA] \subset Y \subset X$. Let γ be any number in Γ_{++}. Because f is twice continuously differentiable, its first and second derivatives are bounded on the domain $Y^{L+1} \times [0, \gamma]$. Since $\sum_{\ell=1}^{L} \ell \, |f_\ell(x, x, \ldots, x; 0)| \leq a$, we have that g and g' are also bounded on Y. Let B be a number that serves as a bound for all of the bounded items. That is, on domain $Y^{L+1} \times [0, \gamma]$, the first and second derivatives of f are all between $-B$ and B. Similarly $|g(x)| \leq B$ and $|g'(x)| \leq B$ for every $x \in Y$. Also, $\left|x^\ell(\Delta) - x^{\ell-1}(\Delta)\right| / \Delta \leq B$ for $\Delta \in (0, \gamma)$ and $\ell \in P_1^L$. Additionally, let

$$S(\Delta) \equiv \max\{k \in P_0^{[T/\Delta]} \mid x^i \in Y \text{ for all } i \leq k\}.$$

Most of the steps in this proof will constrain attention to periods $0, 1, \ldots, S(\Delta)$, where the state is in Y.

We will compare $\{x^k(\Delta)\}_{k=0}^{[T/\Delta]+1}$ with another sequence $\{z^k(\Delta)\}_{k=0}^{[T/\Delta]+1}$ that is defined by the natural discrete-time approximation of y using the known derivative g:

$$z^0(\Delta) = \underline{x} \quad \text{and} \quad z^{k+1}(\Delta) = z^k(\Delta) + g(z^k(\Delta))\Delta \quad \text{for each } k \in P_0^{[T/\Delta]}. \tag{5}$$

We translate this sequence into a step function of continuous time, $\hat{z} \colon [0, T] \times \mathbb{R}_{++} \to X$, by $\hat{z}(t; \Delta) \equiv z^{[t/\Delta]}(\Delta)$.

In the analysis to follow, Δ is always taken to be in $(0, \gamma)$. Because it would be cumbersome in the notation to make explicit the dependence of x^k and z^k on Δ (by always writing "$x^k(\Delta)$" for instance), this dependence is suppressed in most of the expressions below. Also, for any $k \in P_L$, let $\phi^k \equiv (x^k, x^{k-1}, \ldots, x^{k-L})$ denote the vector of states in order from period k back to period $k - L$, and let $\theta^k \equiv (x^k, x^k, \ldots, x^k)$ be the corresponding vector signifying value x^k repeated $L + 1$ times.

The proof proceeds in a series of steps, combining standard methods (Taylor approximations and the like) with a fairly intricate construction in the end. Lemma 1 establishes the first claim, that the initial-value problem has a unique solution, and Lemma 2 establishes that $\hat{z}(\cdot; \Delta)$ converges uniformly to y as $\Delta \to 0^+$. Lemmas 3–5 then identify key properties of $\{x^k(\Delta)\}_{k=0}^{S(\Delta)+1}$ and $\{z^k(\Delta)\}_{k=0}^{S(\Delta)+1}$. Lemma 6 shows that $\{x^k(\Delta)\}_{k=0}^{S(\Delta)+1}$ and $\{z^k(\Delta)\}_{k=0}^{S(\Delta)+1}$ converge uniformly as $\Delta \to 0^+$; the proof uses Lemmas 3 and 4 along with some novel calculations. Finally, we establish that

$S(\Delta) = [T/\Delta]$ for Δ sufficiently small, which implies that the convergence result holds for the entire interval $[0, T]$ in continuous time.

Lemma 1: *There is a unique solution $y\colon [0,T] \to X$ to the initial-value problem $y' = g(y)$ with $y(0) = \underline{x}$.*

Proof of Lemma 1: Existence of a unique solution follows from Picard's Theorem (see, for instance, Süli and Mayers 2003). The requirement's of Picard's Theorem are verified as follows. The relevant interval of states is $\underline{X} \equiv [\underline{x} - TA, \underline{x} + TA]$. Function g is continuous because the derivatives of f are continuous, and $|g|$ is bounded by A. Because g is differentiable and $|g'|$ is bounded by B, it is the case that $|g(x) - g(x')| \leq B|x - x'|$ for all $x, x' \in \underline{X}$ (Lipschitz continuity). Finally, the inequality $TAB \geq A\left(e^{BT} - 1\right)$ holds, regardless of the values of B and T. $\square$

Lemma 2: *$\hat{z}(\cdot;\Delta)$ converges to y uniformly on $[0,T]$ as $\Delta \to 0^+$.*

Proof of Lemma 2: This lemma is a trivial extension of a standard convergence result for Euler's one-step approximation of the solution of an initial-value problem (see, for instance, Süli and Mayers 2003, pages 317–323), and so a proof is not shown here. The requirement that g satisfies a Lipschitz condition follows from g being continuously differential on Y, which is implied by f being twice continuously differentiable on X. The extension here is just to put the result in terms of the step function $\hat{z}$, which uses the fact that $|g| < A$.

$\square$

Lemma 3: *For any given $D > B/(1-a)$, there exists a number $\breve{\Delta} \in (0, \gamma)$ such that $|x^k(\Delta) - x^{k-1}(\Delta)| \leq D\Delta$ for every $\Delta \in (0, \breve{\Delta})$ and every $k \in P_1^{S(\Delta)+1}$. Furthermore, $S(\Delta) \geq [AT/D\Delta]$.*

Proof of Lemma 3: Define constant $B_1 \equiv BL^3D^2 + B(L+1)L^2D + B$ and let $\tilde{\Delta}$ be the number that satisfies $B + \tilde{\Delta}B_1 = D(1-a)$. Note that $\tilde{\Delta}$ is strictly positive because $D > B/(1-a)$. Define sequence $\{b^k(\Delta)\}_{k=1}^{\infty}$ inductively by $b^k(\Delta) \equiv B\Delta$ for each $k = 1, 2, \ldots, L$, and

$$b^{k+1}(\Delta) \equiv ab^k(\Delta) + D(1-a)\Delta$$

for each $k \in P_{L+1}$. Note that, because $a \in [0, 1)$, this sequence is weakly increasing and converges to $D\Delta$. In fact, for $k \geq L$,

$$b^k(\Delta) = D\Delta - (D\Delta - B\Delta)\, a^{k-L}.$$

We will show that $|x^k - x^{k-1}| \leq b^k(\Delta)$ for every $k \in P_1^{S(\Delta)+1}$. This condition holds by assumption for $k \leq L$, so we have to establish the same for $k > L$.

The sequence of states is defined so that $x^{k+1} = f(\phi^k; \Delta)$ for every $k \in P_L^{[T/\Delta]}$. We can use the first-degree Taylor polynomial centered at $(\theta^k; 0)$ to write

$$f(\phi^k; \Delta) = f(\theta^k; 0) + \sum_{\ell=1}^{L} f_\ell(\theta^k; 0)(x^{k-\ell} - x^k) + f_\Delta(\theta^k; 0)\Delta + E^2(k, \Delta),$$

where $E^2(k, \Delta)$ is $1/2$ times the Hessian matrix of f evaluated at some point $(\tilde{\phi}^k; \tilde{\Delta}^k)$ between $(\phi^k; \Delta)$ and $(\theta^k; 0)$, pre- and post-multiplied by the vector $(\phi^k - \theta^k; \Delta)$. Note that there is no term for the derivative of f with respect to its first argument because the first components of ϕ^k and θ^k are both x^k. Noting that $f(\theta^k; 0) = x^k$, we have:

$$x^{k+1} - x^k = \sum_{\ell=1}^{L} f_\ell(\theta^k; 0)(x^{k-\ell} - x^k) + f_\Delta(\theta^k; 0)\Delta + E^2(k, \Delta). \tag{6}$$

An inductive step establishes that $|x^k - x^{k-1}| \leq b^k$ for $\Delta < \tilde{\Delta}$ and every $k \in P_L^{S(\Delta)+1}$. Suppose $k \leq S(\Delta)$ and $|x^{k'} - x^{k'-1}| \leq b^{k'}$ for every $k' \leq k$, as is the case for $k = L$. Using the triangle inequality and that the sequence of bounds is increasing, we then have that

$$\left|x^{k-\ell} - x^k\right| \leq b^k(\Delta) + b^{k-1}(\Delta) + \cdots + b^{k-\ell+1}(\Delta) \leq \ell\, b^k(\Delta).$$

This gives us a bound on the magnitude of the summation term in Equation 6:

$$\left|\sum_{\ell=1}^{L} f_\ell(\theta^k; 0)(x^{k-\ell} - x^k)\right| \leq \sum_{\ell=1}^{L} \ell\, \left|f_\ell(\theta^k; 0)\right| b^k(\Delta) \leq a b^k(\Delta).$$

We can similarly bound the second and third terms on the right side of Equation 6. Since $x^k \in Y$, we have that $|f_\Delta(\theta^k; 0)| \leq B$. Regarding the error term $E^2(k, \Delta)$, because $k \leq S(\Delta)$ and $(\tilde{\phi}^k; \tilde{\Delta}^k)$ is between points in $\{x^0, x^1, \ldots, x^k\}$, we know that $(\tilde{\phi}^k; \tilde{\Delta}^k) \in Y$, where each of the second derivatives of f is bounded by B. Therefore the numbers in the Hessian matrix for $E^2(k, \Delta)$ are all bounded by B and $-B$, implying

$$\left|E^2(i, \Delta)\right| \leq BL\omega(k, \Delta)^2 + B(L+1)L\omega(k, \Delta)\Delta + B\Delta^2,$$

where $\omega(k, \Delta) \equiv \max_{\ell=1,2,\ldots,L} |x^{k-\ell} - x^k|$. Note that $\omega(k, \Delta) \leq Lb^k \leq LD\Delta$. Therefore,

$$\left|E^2(i, \Delta)\right| \leq BL^3D^2\Delta^2 + B(L+1)L^2D\Delta^2 + B\Delta^2 = B_1\Delta^2 < B_1\tilde{\Delta}\Delta.$$

Taking absolute values of the terms in Equation 6 and applying the bound on the magnitudes of each term on the right, we obtain

$$\left|x^{k+1} - x^k\right| \leq ab^k(\Delta) + B\Delta + B_1\tilde{\Delta}\Delta \leq ab^k(\Delta) + D(1-a)\Delta = b^{k+1}(\Delta).$$

Thus, $b^{k+1}(\Delta)$ is a valid bound on $|x^{k+1} - x^k|$, completing the inductive argument.

Because $b^k(\Delta) < D\Delta$ for all k, we have thus shown that $|x^k(\Delta) - x^{k-1}(\Delta)| < D\Delta$ for every $k \in P_1^{S(\Delta)+1}$ and $\Delta \in (0, \tilde{\Delta})$.

On the second claim, recall that $x^0(\Delta)$ converges to $\underline{x}$ as $\Delta \to 0^+$. Let $\breve{\Delta}$ be any number in the interval $(0, \min\{\tilde{\Delta}, \gamma\})$ for which $\Delta < \breve{\Delta}$ implies $|x^0(\Delta) - \underline{x}| < \xi - AT$. By the triangle inequality, we then have $|x^k - \underline{x}| \leq D\Delta k + \xi - AT$, and so $|x^k - \underline{x}| \leq \xi$ for each $k \leq [AT/D\Delta]$, which from the definition of Y implies $S(\Delta) \geq [AT/D\Delta]$. $\square$

For the next lemma and analysis to follow, we will bound the difference between $(x^i(\Delta) - x^{i-1}(\Delta))/\Delta$ and $g(x^k(\Delta))$ for values of i and k between 1 and $S(\Delta)+1$, where k is between $i-1$ and $i+L-1$. To ease notation, define

$$Q(T,\Delta) \equiv \left\{(i,k) \,\middle|\, i \in P_1^{S(\Delta)+1},\ k \in P_0^{S(\Delta)+1},\ \text{and } i-1 \leq k \leq i+L-1\right\}.$$

Then for any $(i,k) \in Q(T,\Delta)$, define

$$m^{i,k}(\Delta) \equiv \frac{x^i(\Delta) - x^{i-1}(\Delta)}{\Delta} - g(x^k(\Delta)).$$

Lemma 4: *There exist a number C, a number $\bar{\Delta} > 0$, and a decreasing function $\kappa\colon (0,\bar{\Delta}) \to \mathbb{R}_+$ such that the following holds for all $\Delta < \bar{\Delta}$. First, $|m^{i,k}(\Delta)| \leq C\Delta$ for all $(i,k) \in Q(T,\Delta)$ satisfying $i > \kappa(\Delta)$. Second, $|m^{i,k}(\Delta)| \leq C$ for all $(i,k) \in Q(T,\Delta)$ satisfying $i \leq \kappa(\Delta)$. Third, $\Delta\kappa(\Delta)$ converges to 0 as $\Delta \to 0^+$.*

Proof of Lemma 4: Let D and $\breve{\Delta}$ be any values that satisfy the claim of Lemma 3 and let us make $\breve{\Delta}$ small enough so that $[AT/D\breve{\Delta}] > 2L$ (as the lemma allows).

For every $\Delta < \breve{\Delta}$ we will construct a bounding sequence $\{b^i(\Delta)\}_{i=1}^{\infty}$, such that for Δ small enough, $|m^{i,k}(\Delta)| \leq b^i(\Delta)$ for all $(i,k) \in Q(T,\Delta)$. Let us first deal with $i \leq L$ and then with $i > L$.

Because $S(\Delta) > 2L$, we know that $x^k \in Y$ and thus $|g(x^k)| \leq B$ for each $k \in P_0^{2L}$. Also, from Lemma 3, we have that $|x^i - x^{i-1}|/\Delta \leq D$ for $i \leq S(\Delta)$. Thus, for $(i,k) \in Q(T,\Delta)$ with $i \leq L$, we have that $|m^{i,k}(\Delta)| \leq B + D$. Let us therefore set $b^i(\Delta) \equiv B + D$ for $i = 1, 2, \ldots, L$.

The bounds for $i > L$ are set using an iterative procedure. To reach the inductive expression, several calculations are required. Consider any $(i,k) \in Q(T,\Delta)$ with $i > L$. By the definition of g and that $x^i = f(\phi^{i-1};\Delta)$, we have

$$m^{i,k}(\Delta) = \frac{f(\phi^{i-1};\Delta) - x^{i-1}}{\Delta} - g(x^k). \tag{7}$$

Let us substitute for $f(\phi^{i-1};\Delta)$ and $g(x^k)$ using Taylor polynomials.

Regarding $f(\phi^{i-1};\Delta)$, we can use the first-degree Taylor polynomial centered at $(\theta^{i-1};0)$ to write

$$f(\phi^{i-1};\Delta) \,=\, f(\theta^{i-1};0) + \sum_{\ell=1}^{L} f_\ell(\theta^{i-1};0)(x^{i-1-\ell} - x^{i-1}) + f_\Delta(\theta^{i-1};0)\Delta + E^2(i,\Delta),$$

where $E^2(i,\Delta)$ has the same form as in the proof of Lemma 3 and satisfies

$$\begin{aligned} |E^2(i,\Delta)| &\le BL\omega(i-1,\Delta)^2 + B(L+1)L\omega(i-1,\Delta)\Delta + B\Delta^2 \\ &= BL\Delta^2 \cdot \left(\frac{\omega(i-1,\Delta)}{\Delta}\right)^2 + B(L+1)L\Delta^2 \cdot \frac{\omega(i-1,\Delta)}{\Delta} + B\Delta^2, \end{aligned}$$

for ω defined as before, so $\omega(i-1,\Delta) \equiv \max_{\ell=1,2,\ldots,L} |x^{i-1-\ell}(\Delta) - x^{i-1}(\Delta)|$. Lemma 3 establishes a uniform bound on $\omega(i-1,\Delta)/\Delta$, namely DL, and hence

$$|E^2(i,\Delta)| \le B_1\Delta^2$$

where B_1 is the same constant defined in the proof of Lemma 3.

Regarding $g(x^k)$, we can use the zero-degree Taylor approximation centered at x^{i-1} to write $g(x^k) = g(x^{i-1}) + E^1(i,k,\Delta)$, where $E^1(i,k,\Delta) = g'(\tilde{x})(x^k - x^{i-1})$ for some number $\tilde{x}$ between x^k and x^{i-1}. Because $(i,k) \in Q(T,\Delta)$, we know that $|x^k - x^{i-1}| \le DL\Delta$ from Lemma 3. Since $|g'|$ on domain Y is bounded by B, it is therefore the case that

$$|E^1(i,k,\Delta)| \le BDL\Delta.$$

Let us next plug into Equation 7 the expressions for $f(\phi^{i-1};\Delta)$ and $g(x^k)$ that we just derived. Using the fact that $f(\theta^{i-1};0) = x^{i-1}$ and $g(x^{i-1}) = f_\Delta(\theta^{i-1};0)/(1 + \sum_{\ell=1}^{L} \ell\, f_\ell(\theta^{i-1};0))$, we obtain:

$$\begin{aligned} m^{i,k}(\Delta) &= \frac{f(\phi^{i-1};\Delta) - x^{i-1}}{\Delta} - g(x^k) \\ &= \sum_{\ell=1}^{L} f_\ell(\theta^{i-1};0)\left(\frac{x^{i-1-\ell} - x^{i-1}}{\Delta}\right) + f_\Delta(\theta^{i-1};0) - g(x^{i-1}) \\ &\qquad + \frac{E^2(i,\Delta)}{\Delta} - E^1(k,i,\Delta) \\ &= \sum_{\ell=1}^{L} f_\ell(\theta^{i-1};0)\left(\frac{x^{i-1-\ell} - x^{i-1}}{\Delta}\right) + g(x^{i-1})\sum_{\ell=1}^{L} \ell f_\ell(\theta^{i-1};0) \\ &\qquad + \frac{E^2(i,\Delta)}{\Delta} - E^1(i,k,\Delta). \end{aligned}$$

Noting that $x^{i-1} - x^{i-1-\ell} = \sum_{j=i-\ell}^{i-1} (x^j - x^{j-1})$, where j takes ℓ values, we can rearrange terms to obtain

$$\begin{aligned} m^{i,k}(\Delta) &= -\sum_{\ell=1}^{L} f_\ell(\theta^{i-1};0) \cdot \sum_{j=i-\ell}^{i-1} \left(\frac{x^j - x^{j-1}}{\Delta} - g(x^{i-1})\right) + \frac{E^2(i,\Delta)}{\Delta} - E^1(i,k,\Delta) \\ &= -\sum_{\ell=1}^{L} f_\ell(\theta^{i-1};0) \cdot \sum_{j=i-\ell}^{i-1} m^{j,\,i-1}(\Delta) + \frac{E^2(i,\Delta)}{\Delta} - E^1(i,k,\Delta). \end{aligned}$$

Taking absolute values and defining $c \equiv B_1 + BDL$ to combine the bounds on the error terms, the last equation implies

$$|m^{i,k}(\Delta)| \leq \sum_{\ell=1}^{L} |f_\ell(\theta^{i-1};0)| \cdot \sum_{j=i-\ell}^{i-1} |m^{j,\,i-1}(\Delta)| + c\Delta.$$

Define $\mu^{i-1}(\Delta) \equiv \max\{|m^{i-L,\,i-1}(\Delta)|, |m^{i-L+1,\,i-1}(\Delta)|, \ldots, |m^{i-1,\,i-1}(\Delta)|\}$. Substituting $\mu^{i-1}(\Delta)$ for $|m^{j,\,i-1}(\Delta)|$ in the previous inequality yields

$$|m^{i,k}(\Delta)| \leq \sum_{\ell=1}^{L} \ell\ |f_\ell(\theta^{i-1};0)|\, \mu^{i-1}(\Delta) + c\Delta,$$

and recalling that $\sum_{\ell=1}^{L} \ell\ |f_\ell(\theta^{i-1};0)| < a < 1$, we conclude that

$$|m^{i,k}(\Delta)| \leq a\, \mu^{i-1}(\Delta) + c\Delta. \tag{8}$$

The inductive step to define $b^i(\Delta)$ for $i > L$ is based on Inequality 8. Recall that we have set $b^1(\Delta) = b^2(\Delta) = \cdots = b^L(\Delta) \equiv B + D$. For every $i \in P_{L+1}$, let us set

$$b^i(\Delta) \equiv ab^{i-L}(\Delta) + c\Delta.$$

Note that $\{b^i(\Delta)\}$ comprises blocks of length L where the value is the same in each block. That is, for each $n \in P_1$, we have

$$b^{nL+1}(\Delta) = b^{nL+2}(\Delta) = \cdots = b^{(n+1)L}(\Delta) \equiv ab^{nL}(\Delta) + c\Delta.$$

Let $\bar{\Delta}$ be any number in the interval $(0, \min\{\check{\Delta}, (B+D)(1-a)/c\})$. For $\Delta < \bar{\Delta}$, the equation $b = ab + c\Delta$ is solved by the value $b = c\Delta/(1-a)$ that is less than $B + C$, implying that the sequence $\{b^i(\Delta)\}$ is weakly decreasing and converges to $c\Delta/(1-a)$. In fact, we have

$$b^{(n+1)L}(\Delta) = \frac{c\Delta}{1-a} + \left(B + D - \frac{c\Delta}{1-a}\right) a^n$$

for each $n \in P$.

We next have an inductive step to validate that $\{b^i(\Delta)\}$ provides the desired bounds. Consider any $i \in P_{L+1}^{S(\Delta)+1}$ for which $|m^{i',k'}(\Delta)| \leq b^{i'}(\Delta)$ for all $(i',k') \in Q(T,\Delta)$ such that $i' < i$. (We have already shown that the presumption holds for $i = L+1$.) That $\{b^{i'}(\Delta)\}$ is weakly decreasing implies that $\mu^{i-1}(\Delta) \leq b^{i-L}(\Delta)$. From Inequality 8, this further implies that $|m^{i,k}(\Delta)| \leq ab^{i-L}(\Delta) + c\Delta = b^i(\Delta)$ for all $k \in P$ satisfying $(i,k) \in Q(T, |De)$, validating the bound for i.

So we have proven that $|m^{i,k}(\Delta)| \leq b^i(\Delta)$ for all $(i,k) \in Q(T,\Delta)$. The final step of the proof is to show that $\{b^i(\Delta)\}_{i=1}^{\infty}$ has the three stated properties for Δ close enough to 0. The idea is to characterize at which period $b^i(\Delta)$ crosses a threshold multiple of its limit. Letting the multiple be 2, define function β so that, for every Δ,

$$\frac{c\Delta}{1-a} + \left(B + D - \frac{c\Delta}{1-a}\right) a^{\beta(\Delta)} = \frac{2c\Delta}{1-a}.$$

That is, letting $\rho(\Delta) \equiv \frac{c\Delta}{1-a} / \left(B + D - \frac{c\Delta}{1-a}\right)$, we have $a^{\beta(\Delta)} = \rho(\Delta)$, and therefore $\beta(\Delta) = \ln \rho(\Delta) / \ln a$. Then $b^i(\Delta) \leq 2c\Delta/(1-a)$ for every $i > (\beta(\Delta)+1)L$.

Define $\kappa(\Delta) \equiv (\beta(\Delta)+1)L$ and let $C = \max\{2c/(1-a), B+D\}$. We showed in the previous paragraph that $b^i(\Delta) < C\Delta$ for every $i > \kappa(\Delta)$, proving the first claim of Lemma 4. The second claim follows from the fact that $b^1(\Delta) = B + D$ and $\{b^i(\Delta)\}$ is decreasing. Straightforward calculations show that $\lim_{\Delta \to 0^+} \Delta \ln \rho(\Delta) = 0$ (using L'Hôpital's rule), which proves the third claim. $\square$

Lemma 5: *$|z^k(\Delta) - \underline{x}| \leq kA\Delta$ for each $k \in P_0^{[T/\Delta]+1}$.*

Proof of Lemma 5: This follows directly from Expression 5 and that $|g|$ is bounded above by A on domain $[\underline{x} - TA, \underline{x} + TA]$. $\square$

With the previous lemmas in hand, the next lemma goes most of the way toward proving Theorem 1 by constructing a uniform upper bound on $\left|x^k(\Delta) - z^k(\Delta)\right|$ for each $k \in P_0^{S(\Delta)+1}$.

Lemma 6: *For any given $\varepsilon > 0$ there exists a number $\eta > 0$ such that $\Delta \in (0, \eta)$ implies $\left|x^k(\Delta) - z^k(\Delta)\right| < \varepsilon$ for all $k \in P_0^{S(\Delta)+1}$.*

Proof of Lemma 6: Let D and $\bar{\Delta}$ be any values that satisfy the claims of Lemmas 3 and 4 (with $\bar{\Delta} \leq \breve{\Delta}$). Let us start by formulating a bound for $k \in P_1^{S(\Delta)+1}$ with $k \leq [\kappa(\Delta)] + L$, where $\kappa(\Delta)$ is defined by Lemma 4. We'll then assess all larger values of k using an inductive argument.

Invoking Lemma 3 for $k \leq [\kappa(\Delta)] + L$, we have

$$|x^k - x^0| \leq ([\kappa(\Delta)] + L)D\Delta.$$

Likewise, Lemma 5 implies that

$$|z^k - \underline{x}| \leq ([\kappa(\Delta)] + L)B\Delta.$$

From Lemma 4 we know that $\Delta\kappa(\Delta)$ converges to 0, implying that the right sides of both of these inequalities converge to 0 as $\Delta \to 0^+$. Recall as well that $x^0(\Delta)$ converges to $\underline{x}$. These facts together imply that there exists a function $\lambda \colon \mathbb{R} \to \mathbb{R}$ such that, for $\Delta < \bar{\Delta}$, it is the case that $|x^k(\Delta) - z^k(\Delta)| \leq \lambda(\Delta)$ for every $k \leq [\kappa(\Delta)] + L$, and $\lambda(\Delta)$ converges to 0 as $\Delta \to 0^+$.

Next let us look at values of k greater than $[\kappa(\Delta)] + L$. A key step is figuring how to write $x^{k+1} - z^{k+1}$ as a function of $x^k - z^k$. Using Equations 5 and 2, we have

$$x^{k+1}(\Delta) - z^{k+1}(\Delta) = f(\phi^k; \Delta) - z^k(\Delta) - g(z^k(\Delta))\Delta.$$

Substituting for $f(\phi^k;\Delta)$ using the first-degree Taylor polynomial centered at $(\theta^k;0)$, and using $f(\theta^k;0)=x^k$ and the identity

$$g(x) \equiv \frac{f_\Delta(x,x,\ldots,x;0)}{1+\sum_{\ell=1}^{L} \ell\, f_\ell(x,x,\ldots,x;0)}$$

to substitute terms, we obtain

$$\begin{aligned}
x^{k+1}-z^{k+1} &= f(\theta^k;0)+\sum_{\ell=1}^{L} f_\ell(\theta^k;0)(x^{k-\ell}-x^k)+f_\Delta(\theta^k;0)\Delta+E^2(k,\Delta)\\
&\qquad -z^k-g(z^k)\Delta\\
&= x^k-z^k+\sum_{\ell=1}^{L} f_\ell(\theta^k;0)(x^{k-\ell}-x^k)+f_\Delta(\theta^k;0)\Delta-g(x^k)\Delta\\
&\qquad +\left[g(x^k)-g(z^k)\right]\Delta+E^2(k,\Delta)\\
&= x^k-z^k+\sum_{\ell=1}^{L} f_\ell(\theta^k;0)(x^{k-\ell}-x^k)+g(x^k)\Delta\sum_{\ell=1}^{L}\ell f_\ell(\theta^k;0)\\
&\qquad +\left[g(x^k)-g(z^k)\right]\Delta+E^2(k,\Delta),
\end{aligned}$$

where the error term $E^2(k,\Delta)$ is characterized as in the proofs of the previous two lemmas; that is, for the constant B_1 defined before,

$$|E^2(k,\Delta)| \le B_1\Delta^2.$$

Let us deal with the term $x^{k-\ell}-x^k$ also along the lines of the proof of Lemma 4. Noting that $x^k-x^{k-\ell}=\sum_{i=k-\ell+1}^{k}(x^i-x^{i-1})$, where i takes ℓ values, we rearrange terms to obtain

$$\begin{aligned}
x^{k+1}-z^{k+1} &= x^k-z^k-\Delta\sum_{\ell=1}^{L} f_\ell(\theta^k;0)\cdot\sum_{i=k-\ell+1}^{k}\left(\frac{x^i-x^{i-1}}{\Delta}-g(x^k)\right)\\
&\qquad +\left[g(x^k)-g(z^k)\right]\Delta+E^2(k,\Delta)\\
&= x^k-z^k-\Delta\sum_{\ell=1}^{L} f_\ell(\theta^k;0)\cdot\sum_{i=k-\ell+1}^{k} m^{i,k}(\Delta)\\
&\qquad +\left[g(x^k)-g(z^k)\right]\Delta+E^2(k,\Delta).
\end{aligned}$$

Next use the zero-degree Taylor approximation for g, centered at z^k, to write

$$g(x^k)=g(z^k)+g'(\tilde{x}^k)(x^k-z^k),$$

where $\tilde{x}^k$ is a number between x^k and z^k. Because x^k and z^k are in Y, so is $\tilde{x}^k$ and therefore $|g'(\tilde{x}^k)|\le B$. Substituting for $g(x^k)$ in the expression for $x^{k+1}-z^{k+1}$, using the definition of $m^{i,k}$, and rearranging terms, we obtain

$$x^{k+1} - z^{k+1} = (x^k - z^k)\left(1 + g'(\tilde{x}^k)\Delta\right) - \Delta \sum_{\ell=1}^{L} f_\ell(\theta^k; 0) \cdot \sum_{i=k-\ell+1}^{k} m^{i,k}(\Delta) + E^2(k, \Delta). \quad (9)$$

For $k \geq \kappa(\Delta) + L$, it is the case that $i > \kappa(\Delta)$ in the summation of $m^{i,k}$ on the right and so, from Lemma 4, $|m^{i,k}(\Delta)| \leq C\Delta$. In this summation, i takes ℓ values, and therefore $|\sum_{i=k-\ell+1}^{k} m^{i,k}(\Delta)| \leq \ell\, C\Delta$, which implies:

$$\left| \Delta \sum_{\ell=1}^{L} f_\ell(\theta^k; 0) \cdot \sum_{i=k-\ell+1}^{k} m^{i,k}(\Delta) \right| \leq C\Delta^2 \sum_{\ell=1}^{L} \ell |f_\ell(\theta^k; 0)| \leq aC\Delta^2.$$

Taking absolute values of all components of Equation 9 and using the bounds on $|E^2(k, \Delta)|$ and $|g'(\tilde{x}^k)|$, we obtain:

$$|x^{k+1} - z^{k+1}| \leq |x^k - z^k|\,(1 + B\Delta) + aC\Delta^2 + B_1\Delta^2. \quad (10)$$

Rearranging terms yields:

$$\left|x^{k+1} - z^{k+1}\right| + \frac{aC + B_1}{B}\Delta \leq (1 + B\Delta)\left(\left|x^k - z^k\right| + \frac{aC + B_1}{B}\Delta\right).$$

By induction, we have for $\kappa(\Delta) + L \leq k \leq S(\Delta) + 1$,

$$\begin{aligned} \left|x^{k+1} - z^{k+1}\right| + \frac{aC + B_1}{B}\Delta &\leq (1 + B\Delta)^{k-(\kappa(\Delta)+L)}\left(\left|x^{\kappa(\Delta)+L} - z^{\kappa(\Delta)+L}\right| + \frac{aC + B_1}{B}\Delta\right) \\ &\leq (1 + B\Delta)^{[T/\Delta]+1}\left(\left|x^{\kappa(\Delta)+L} - z^{\kappa(\Delta)+L}\right| + \frac{aC + B_1}{B}\Delta\right). \end{aligned}$$

Since $\left|x^{\kappa(\Delta)+L} - z^{\kappa(\Delta)+L}\right| \leq \lambda(\Delta)$, we know that

$$\left|x^k - z^k\right| \leq (1 + B\Delta)^{[T/\Delta]+1}\lambda(\Delta) + \left((1 + B\Delta)^{[T/\Delta]+1} - 1\right)\left(\frac{aC + B_1}{B}\Delta\right).$$

The right-hand side is a uniform upper bound that does not depend on k, and it converges to zero as $\Delta \to 0$ since $(1 + B\Delta)^{[T/\Delta]+1} \to e^{BT}$ and $\lambda(\Delta) \to 0$.

□

Because step functions $\hat{x}$ and $\hat{z}$ are defined in the same manner relative to sequences $\{z^k(\Delta)\}_{k=0}^{[T/\Delta]+1}$ and $\{x^k(\Delta)\}_{k=0}^{[T/\Delta]+1}$, Lemma 6 establishes that $\hat{x}$ and $\hat{z}$ converge uniformly on $[0, \bar{t}]$ for every number $\bar{t} > 0$ for which $S(\Delta)\Delta$ weakly exceeds $\bar{t}$ for sufficiently small Δ. From Lemma 3, we know this is the case for $\bar{t} = AT/D$.

In fact, it must also be the case for $\bar{t} = T$. To show this, let us presume otherwise and we shall find a contradiction. Suppose $\liminf_{\Delta\to 0^+} S(\Delta)\Delta < T$. Then there exists $\varepsilon_0 > 0$ such that for every $\hat{\Delta} > 0$ there is a number $\Delta \in (0, \hat{\Delta})$ for which $S(\Delta)\Delta < T - \varepsilon_0$. Using Lemma 5, we thus have $\left|z^{S(\Delta)} - \underline{x}\right| \leq AS(\Delta)\Delta < AT - A\varepsilon_0$. By Lemmas 3 and 5 we know that both $\left|x^{S(\Delta)+1} - x^{S(\Delta)}\right|$ and $\left|z^{S(\Delta)} - x^{S(\Delta)}\right|$ can be made arbitrarily small by taking $\hat{\Delta}$ sufficiently small. Therefore, the triangle inequality implies

$$\left|x^{S(\Delta)+1} - \underline{x}\right| \leq \left|x^{S(\Delta)+1} - x^{S(\Delta)}\right| + \left|z^{S(\Delta)} - x^{S(\Delta)}\right| + \left|z^{S(\Delta)} - \underline{x}\right| < AT$$

for sufficiently small $\hat{\Delta}$. This implies that $x^{S(\Delta)+1}(\Delta) \in [\underline{x} - AT, \underline{x} + AT]$ for some values of Δ, contradicting the definition of $S(\Delta)$.

We have proved that $\hat{x}$ and $\hat{z}$ converge uniformly on $[0, T]$. Lemma 2 established the uniform convergence of $\hat{z}$ to y on $[0, T]$, and therefore $\hat{x}$ also converges uniformly to y on $[0, T]$.

4.2 Proof of the Second Theorem

This section contains a proof of Theorem 2. The steps of this proof are straightforward modifications of a small subset of steps shown for Theorem 1, so I informally summarize them here.

Lemmas 1 and 2 extend to an n-component system of differential equations by simply replacing the absolute value sign with the Euclidean norm (see, for instance, Süli and Mayers 2003 and Coddington and Levinson 1955). So what is left is to prove that $\{x^k(\Delta)\}_{k=0}^{S(\Delta)+1}$ and $\{z^k(\Delta)\}_{k=0}^{S(\Delta)+1}$ converge. The complicated arguments used in the proof of the main theorem are not needed for the present proof because they had mainly to do with the lagged states ($L > 0$). In place of Lemmas 3-6, we can proceed with a direct comparison of the two sequences and construct bounds via induction.

As in the proof of the main theorem, we can find closed convex set $Y \subset X$ that contains the ball of radius $TA + \varepsilon$ around $\underline{x}$ for some $\varepsilon > 0$. Let $-B$ and B bound all components of the first and second derivatives of f and the first derivatives of g on the set Y. The analysis to follow is restricted to values of k such that $x^0, x^1, \ldots, x^k \in Y$.

From the sequence definitions, we have

$$x^{k+1}(\Delta) - z^{k+1}(\Delta) \quad = \quad f(x^k(\Delta); \Delta) - z^k(\Delta) - g(z^k(\Delta))\Delta.$$

Substituting for the first term on the right side using the first-degree Taylor polynomial centered at $(x^k(\Delta), 0)$, and using the identity $g(x) = f_\Delta(x; 0)$ and that $f(x, 0) = x$, we obtain

$$\begin{aligned}
x^{k+1} - z^{k+1} &= f(x^k; 0) + f_\Delta(x^k; 0)\Delta + E^2(k, \Delta) - z^k(\Delta) - g(z^k)\Delta \\
&= x^k - z^k + f_\Delta(x^k; 0)\Delta - g(z^k)\Delta + E^2(k, \Delta) \\
&= x^k - z^k + \left[g(x^k) - g(z^k)\right]\Delta + E^2(k, \Delta),
\end{aligned}$$

where the error term satisfies $\|E^2(k, \Delta)\| \leq B\Delta^2$.

Let $G(x)$ denote the matrix of first partial derivatives of g evaluated at x, and let $H^i(x)$ denote the Hessian matrix of g^i. We can use the first-degree Taylor approximation for g, centered at $z^k(\Delta)$, to write

$$g(x^k) = g(z^k) + G(z^k)(x^k - z^k) + \tilde{E}^2(k, \Delta), \tag{11}$$

where $\tilde{E}^2(k,\Delta)$ is a vector whose ith component is equal to $(x^k - z^k)' H^i(\hat{x}^i)(x^k - z^k)/2$ for some $\hat{x}^i$ between x^k and z^k. We thus have $\|\tilde{E}^2(k,\Delta)\| \leq B\|x^k - z^k\|^2$. Using Equation 11 to substitute for $g(x^k)$ in the expression for $x^{k+1} - z^{k+1}$, we obtain

$$x^{k+1} - z^{k+1} = x^k - z^k + G(z^k)\left(x^k - z^k\right)\Delta + \tilde{E}^2(k,\Delta)\Delta + E^2(k,\Delta).$$

Using the bounds on $\tilde{E}^2$ and E^2, and noting that $\|G(z^k)\left(x^k - z^k\right)\Delta\| \leq B\Delta\|x^k - z^k\|$, we have

$$\|x^{k+1} - z^{k+1}\| \leq \|x^k - z^k\| + 2B\Delta\|x^k - z^k\| + B\Delta^2. \tag{12}$$

Consider $\Delta < 1$. We next construct an upper bound on $\|x^k - z^k\|$, denoted $b^k(\Delta)$, for each $k = 1, 2, \ldots$ for which x^k and z^k remain in Y. Set $b^1(\Delta) = B\Delta$, which is clearly an upper bound on $\|x^1 - z^1\|$ due to Inequality 12, that $\Delta \leq 1$, and because $x^0 = z^0$. By letting Δ be small enough, the upper bounds $b^2(\Delta), b^3(\Delta), \ldots$ will be constructed to ensure they are all weakly greater than $B\Delta$ and weakly less than 1. With reference to the right side of Inequality 12, this means that $b^k(\Delta) + 2B\Delta b^k(\Delta) + B\Delta^2 \leq b^k(\Delta)(1 + 3B\Delta)$. Inductively define $b^2(\Delta), b^3(\Delta), \ldots, b^{[T/\Delta]}(\Delta)$ by setting

$$b^{k+1}(\Delta) = b^k(\Delta)(1 + 3B\Delta). \tag{13}$$

Clearly the sequence is increasing, so $b^k(\Delta) \geq B\Delta$ for each k. Recall that we have set $b^1(\Delta) = B\Delta$, and so we can use Equation 13 to explicitly solve for $b^{[T/\Delta]}(\Delta)$:

$$b^{[T/\Delta]}(\Delta) = B\Delta(1 + 3B\Delta)^{[T/\Delta]}.$$

From the definition of the natural number e, the right side is less than $B\Delta e^{3BT}$. For Δ small enough, this number is below 1 and our presumptions hold to ensure that the bounds are valid.

As $\Delta \to 0$, the bounds are all smaller than $b^{[T/\Delta]}(\Delta)$ and thus converge uniformly to 0, and so $\{x^k(\Delta)\}$ and $\{z^k(\Delta)\}$ converge uniformly for each $k = 1, 2, \ldots$ for which x^k and z^k remain in Y. The argument made at the end of the proof of the main theorem applies here as well to establish that this is the case for all $k \leq [T/\Delta]$ provided Δ is small enough.

References

Acemoglu, Daron and James Robinson, "The Emergence of Weak, Despotic and Inclusive States," NBER Working Paper (2018).

Coddington, Earl A. and Norman Levinson, *Theory of Ordinary Differential Equations*, Tata McGraw-Hill Education, 1955.

Hua, Xiameng, and Joel Watson, “Starting Small and Renegotiation in Discrete-Time Relationships with a Continuum of Types,” manuscript (2020).

Solow, Robert M., “A Contribution to the Theory of Economic Growth,” *Quarterly Journal of Economics* 70 (1, 1956): 65–94.

Süli, Endre and David F. Mayers, *An Introduction to Numerical Analysis*, Cambridge: Cambridge University Press, 2003.

Swan, Trevor W., “Economic Growth and Capital Accumulation,” *Economic Record* 32 (2, 1956): 334–361.

Watson, Joel, “Starting Small and Renegotiation,” *Journal of Economic Theory* 85 (1999, 1): 52–90.

Watson, Joel, “Starting Small and Commitment,” *Games and Economic Behavior* 38 (2002, 1): 176–199.