Error Analysis in Dynamic Models I.

by Adrian Peralta-Alva

Federal Reserve Bank of Saint Louis

September 9, 2011

Introduction

- First we study models for which established theory gives foundations to numerical methods.
- Study errors in the policy function, in simulations from such policy, and implications for estimation.
- Roughly speaking, models where equilibrium is Markovian and continuous have the desired properties.
- Monotonicity or contraction properties allow sharper results (error bounds).
- Models with heterogeneous agents and frictions may be problematic (the topic of our next lecture).

Numerical Dynamic Programming

Based on Santos, Manuel S., 1999. "Numerical Solution of Dynamic Economic Models," In: John B. Taylor and Michael Woodford, Editor(s), Handbook of Macroeconomics, Elsevier, Volume 1, Part 1, Pages 311-386

Numerical Dynamic Programming

- Basic idea is restrict the set of functions to a finite-dimensional domain, which can be captured by a finite number of instructions.
- These functions will be defined over a compact domain S via piecewise affine interpolation. Different interpolation schemes can be used.

Numerical Dynamic Programming

- We define a grid as finite collection of simplices (in R^l is the convex combination of $l+1$ points, in R an interval) $\{S^j\}$ such that $\cup S^j = S$, *and* $int(S^i) \cap int(S^j) = \emptyset$ for all $i \neq j$.
- A generic vertex is (k^j) and the grid size is $h = max_j diam(S^j)$.
- The space of functions we will employ, is

$$\mathscr{W}^h = \{V^h : S \to R \mid V^h \text{ is bdd, continuous, and}$$

$$DV^h \text{ is constant in } int(S^j)\}.$$

Algorithm

- For a given grid $\{S^j\}$ with mesh size h the following is the value function iteration algorithm
  1. Initial step: select an accuracy level $TOLW$ and an initial guess $W_0^h \in \mathscr{W}^h$.
  2. Operator $T^h W_n^h$ can be defined at vertex point k^j by

$$W_{n+1}^h(k^j) = \max_{k'} v(k^j, k') + \beta W_n(k')$$

$$s.t.\ (k^j, k') \text{ feasible, and } k^j, k' \in S.$$

  3. End of iteration: If $||W_{n+1} - W_n|| \leq TOLW$ stop; else, increment n by 1 and return to step 2

Assumptions and implications

- The state space is compact, and the feasible set for (k, k') is convex.
- The "indirect utility" v is continuous, and C^2 with bounded derivatives. Moreover, it satisfies a strong form of concavity: There exists a constant $\eta > 0$ such that $v(k, k') + \frac{1}{2}\eta||k'||^2$ is concave on k, k'
- Optimal paths are interior
- Theorem 3.1 and Corollary 3.2: These assumptions imply the true value function is C^2, and the policy function C^1.

Properties of the Numerical Algorithm

- Lemma 4.1. Under our assumptions, the numerical algorithm has a unique limit $W^h \in \mathscr{W}^h$. Further, this limit can be approached to arbitrary precision in a finite number of iterations (that is, for any $\varepsilon > 0$ there is $N < \infty$: $sup_k |W^h_{\tilde{n}} - W^h| < \varepsilon$ for all $\tilde{n} > N$).
- Lemma 4.2. Let W be the true value function (i.e., the limit of the value function iteration operator without discretizing), and γ a bound on its curvature. Then, $||TW - T^h W|| \leq \frac{\gamma}{2} h^2$.
- Theorem 4.3. Let W be the true value function and W^h be the limit of the numerical algorithm above. Then, under our assumptions

$$||W - W^h|| \leq \frac{M}{1-\beta} h^2$$

Where M is a constant that depends on model primitives.

Intuition behind these results

Key for Theorem 4.3: If T, T^h denote the value function iteration, and its discretized counterpart above, then
$||W - W^h|| = ||TW - T^hW^h|| \leq ||TW - T^hW|| + ||T^hW - T^hW^h||$

$$\leq ||TW - T^hW|| + \beta||W - W^h||$$

Hence $||W - W^h|| \leq \frac{||TW - T^hW||}{1-\beta} \leq \frac{\gamma h^2}{2(1-\beta)}$, due to Lemma 4.2.

Modifications to speed up the algorithm

- Value function iteration is reliable, but it is very slow. A lot of iterations are needed, in general, to achieve $||W^h_{n+1} - W^h_n||$ low enough.
- The number of iterations required is a function of how far the initial condition W^h_0 is from the actual fixed point of the operator. Since the true solution is "close" to the fixed point of the operator for any small h, the closer the initial guess is to the true solution, the closer we will converge.

Modifications to speed up the algorithm

- Multigrid methods
- The idea is simple: solve the model under a relatively coarse grid with mesh size h_0, which results in a fixed point W^{h_0}. This function can then be used as initial condition for the value function iteration with a finer mesh size $h_1 < h_0$.

Multigrid

Table 6
Example 7.1. Computational method: dynamic programming algorithm with linear interpolation[a]

Vertex points	Mesh size	Iterations	CPU time	Max. error in g	Max. error in W
100	10^{-1}	460	18.71	4.77×10^{-2}	1.980×10^{-1}
1000	10^{-2}	920	378.35	5.57×10^{-3}	1.949×10^{-3}
10000	10^{-3}	1379	5367.44	5.97×10^{-4}	1.900×10^{-5}

[a] Parameter values: $\beta = 0.99$, $\lambda = \frac{1}{3}$, $A = 10$, $\alpha = 0.34$ and $\delta = 1$.

Table 7
Example 7.1. Computational method: multigrid with linear interpolation[a]

Vertex points	Mesh size	Iterations	CPU time
1000	10^{-2}	460	210.67
10000	10^{-3}	932	1802.19

[a] Parameter values: $\beta = 0.99$, $\lambda = \frac{1}{3}$, $A = 10$, $\alpha = 0.34$ and $\delta = 1$.

Accuracy of Numerical Solutions Using the Euler Equations

Based on Santos, Manuel S., 2000. "Accuracy of Numerical Solutions Using the Euler Equation Residuals," Econometrica, Vol. 68, No. 6, pp. 1377-1402.

Euler equations

- The Euler equation of the one sector growth model is

$$\frac{1}{Ak_t^{\alpha}+(1-\delta)k_t-k_{t+1}}=\beta\frac{\alpha Ak_{t+1}^{\alpha-1}+1-\delta}{Ak_{t+1}^{\alpha}+(1-\delta)k_{t+1}-k_{t+2}}$$

- Further, because of the principle of optimality we know there is a unique $g(k)$ such that the sequence $\{k_t\}$, recursively generated by $k_{t+1}=g(k_t)$, given k_0, solves the planner problem.

Euler equation residuals

- Suppose we have an arbitrary $\tilde{g}(k)$ that satisfies, at all k,

$$\frac{1}{Ak^{\alpha}+(1-\delta)k-\tilde{g}(k)} \approx \beta \frac{\alpha A(\tilde{g}(k))^{\alpha-1}+1-\delta}{A(\tilde{g}(k))^{\alpha}+(1-\delta)\tilde{g}(k)-\tilde{g}(\tilde{g}(k))}$$

- Can we say that $\tilde{g}(k)$ is a good approximation to the true solution of the model?

Intuition from static problems

LEMMA 2.1 (Santos ECTA 2000): Assume that $F : R^l \to R$ is a C^2 mapping. Let $DF(x)$ be the derivative of function F at point x. Assume that F is concave in the following strong sense: There is a constant $\eta > 0$ such that for all x in R^l the function $F(x) + (\eta/2)||x||^2$ is concave. Let $x^* = \arg\max F(x)$. Then $||DF(x)|| \leq \varepsilon$ implies $||x - x^*|| \leq (1/\eta)\varepsilon$ and $||F(x) - F(x^*)|| \leq (1/\eta)\varepsilon^2$.

Proof: for $F : R \to R$. A first order Taylor approximation of F' around x^* yields

$$F'(x) = F'(x^*) + F''(s)(x - x^*)$$

But the first derivative vanishes at the maximizer x^*, and η is a lower bound for the curvature of F. Hence $\varepsilon \geq |F'(x)| \geq \eta|x - x^*|$ together imply $|x - x^*| \leq \varepsilon/\eta$. Finally, by the concavity of F, $F(x) - F(x^*) \leq F'(x)(x - x^*)$, but $|F'(x)| \leq \varepsilon$ and $|x - x^*| \leq \varepsilon/\eta$ and thus $|F(x) - F(x^*)| \leq (1/\eta)\varepsilon^2$.

Dynamic version

- Assumption 1: Endogenous predetermined variables lie in a compact set; the feasible set is convex
- Assumption 2: Value function is C^2 and strongly concave
- Assumption 3: Equilibrium paths lie in the interior of the feasible set for every initial condition

Dynamic version

- To simplify the presentation I focus here on the deterministic version of the model with one state variable, but the results generalize to stochastic models and several states
- The necessary conditions for a path $\{k_t\}$ to be optimal can be written as

$$v_1(k_t, k_{t+1}) + \beta v_2(k_{t+1}, k_{t+2}) = 0,$$

where v is the "indirect" utility (once we wrote everything in terms of the endogenous states), and v_i denotes the partial derivative of this function with respect to its $i-th$ argument

- Define

$$\varepsilon = \max_k |v_1(k, \hat{g}(k)) + \beta v_2(\hat{g}(k), \hat{g}(\hat{g}(k)))|$$

Dynamic version

- Suppose $|g - \hat{g}| \leq \delta$, then if the difference between orbits can be bounded, then the discounted lifetime utility under g, W, and under $\hat{g}$, $W_{\hat{g}}$ satisfy $|W - W_{\hat{g}}| \leq H\varepsilon\delta/(1-\beta)$
- Suppose $|W - W_{\hat{g}}| \leq \gamma$, then $|g - \hat{g}| \leq (2\gamma/\eta)^{1/2}$

Intuition for the proofs

$$W(k) - W_{\hat{g}}(k) = \sum \beta^t (v(k_t, k_{t+1}) - v(\widehat{k}_t, \hat{k}_{t+1})) \leq$$
$$\sum \beta^t (v_1(\widehat{k}_t, \hat{k}_{t+1})(k_t - \hat{k}_t) + v_2(\widehat{k}_t, \hat{k}_{t+1})(k_{t+1} - \hat{k}_{t+1}) =$$
$$v_2(\widehat{k}_0, \hat{k}_1)(k_1 - \hat{k}_1) + \beta v_1(\widehat{k}_1, \hat{k}_2)(k_1 - \hat{k}_1) + ... =$$
$$\sum \beta^t (v_2(\widehat{k}_t, \hat{k}_{t+1}) + \beta v_1(\widehat{k}_{t+1}, \hat{k}_{t+2}))(k_t - \hat{k}_t) \leq$$
$$\delta \sum \beta^t (k_t - \hat{k}_t).$$

If the distance between approximate and actual solution, $|k_t - \hat{k}_t|$, is bounded for all t then we are done.

Intuition for the proofs

Let $f(k,k') = v(k,k') + \beta W(k')$, and $k^* = g(k)$, $\hat{k} = \hat{g}(k)$. The assumption $|W - W_{\hat{g}}| \leq \gamma$, implies $f(k,k^*) - f(k,\hat{k}) \leq \gamma$ Take a second order expansion of f on $\hat{k}$ at point (k,k^*) :

$$f(k,k^*) = f(k,\hat{k}) + f_2(k,k^*)(k-k^*) + \frac{1}{2} f_{22}(k,s)(s-k^*)^2.$$

Of course, $f_2(k,k^*) = 0$ and given curvature bounds

$$\gamma \geq f(k,k^*) - f(k,\hat{k}) \geq \frac{\eta}{2}|k^* - \hat{k}|$$

Main results v1

- Assumption 4: For all $\delta > 0$ there is $H > 0$ such that if $||g - \hat{g}|| \leq \delta$ then $||k_t - \hat{k}_t|| \leq H\delta$ for all $t > 1$.
- Theorem 3.3: Let ε be the maximum Euler equation residual of function $\hat{g}$. Then, under Assumptions 1-4 we have

$$||W - W_{\hat{g}}|| \leq \frac{2H^2\varepsilon^2}{\eta(1-\beta)^2} \qquad ||g - \hat{g}|| \leq \frac{2H\varepsilon}{\eta(1-\beta)}.$$

Main results v2

- Let $L = \max_k |v_{11}(k, g(k))|$
- Theorem 3.5: Let ε be the maximum Euler equation residual of function $\hat{g}$. Then, under Assumptions 1-3 we have

$$||W - W_{\hat{g}}|| \leq \frac{2}{\eta(1/\sqrt{\beta} - 1)^2(1 - \sqrt{\beta})^2} \frac{L}{\eta} \varepsilon^2$$

$$||g - \hat{g}|| \leq \frac{2}{\eta(1/\sqrt{\beta} - 1)(1 - \sqrt{\beta})} \left(\frac{L}{\eta}\right)^{1/2} \varepsilon.$$

Conclusions

- Euler equations can be easily computed for any arbitrary policy function.
- Under standard regularity conditions, the accuracy of approximation of a policy function is proportional to the magnitude of its Euler equation.
- The constant of proportion depends on primitives like the discount factor, the curvature of the utility function, and the curvature of the value function.

Accuracy of Simulations for Stochastic Dynamic Models

Based on Peralta-Alva, Adrian and M. S. Santos, 2005. "Accuracy of Simulations for Stochastic Dynamic Models," Econometrica, 73, 1939-1976

Overview

- Stochastic dynamic models may have a recursive structure. Hence, the state (endogenous) evolves according to

$$k_{t+1} = g(k_t, z_t)$$

  where g is a time invariant policy function, and z_t is an exogenously given stochastic process.
- Computers may be used to simulate sequences of shocks $\{z_t\}$, and thus equilibrium time series for each state history $z^t = z_0, z_1, ..., z_t$, $\{k(z^t), GDP(z^t), c(z^t), inv(z^t)\}_{t \geq 0}$.
- What are the statistical properties one can derive from a model? why? how do we compute them?

The problem

- More important, since stochastic dynamic models can only be numerically approximated, the best we can hope for is to obtain an approximate policy function $\tilde{g}$ and be able to generate approximate time series

$$\tilde{k}_{t+1} = \tilde{g}(\tilde{k}_t, z_t)$$

- Surprisingly, very little work has been devoted to study conditions under which approximation errors in $\tilde{g}$ do not cumulate through time and result in biased simulated statistics.

Framework of analysis

Dynamical system:

$$\begin{aligned} z_{n+1} &= \Psi(z_n, \varepsilon_{n+1}) \\ k_{n+1} &= g(z_n, k_n, \varepsilon_{n+1}), \qquad n = 0, 1, 2, \cdots . \end{aligned} \tag{2.1}$$

where z is a finite vector of exogenous shocks in Z in Euclidean space $\subset R^m$. It evolves according to a function Ψ and an *iid* shock ε in a set of "events" E. The distribution of the shock ε is given by a probability measure Q defined on a measurable space $(E, \mathbb{E})$. k lists endogenous state variables in $K \subset R^l$.
$s = (z, k)$ is a generic vector in $S = Z \times K$.
For expository purposes, we summarize (2.1) as:

$$s_{n+1} = \varphi(s_n, \varepsilon_{n+1}), \qquad n = 0, 1, 2, \cdots . \tag{2.2}$$

Stationary distributions

Stochastic systems can generate very complex dynamics. It is useful to define the transition probability function

$$P(s, A) = Q(\{\varepsilon | \varphi(s, \varepsilon) \in A\}). \tag{2.4}$$

For any given initial condition μ_0 on $\mathbb{S}$, the evolution of future probabilities, $\{\mu_n\}$, can be specified by the following operator T^* that takes the space of probabilities on $\mathbb{S}$ into itself

$$\mu_{n+1}(A) = (T^*\mu_n)(A) = \int P(s, A)\mu_n(ds), \tag{2.5}$$

for all A in $\mathbb{S}$ and $n \geq 0$. An invariant probability measure or invariant distribution μ^* is a fixed point of operator T^*, i.e., $\mu^* = T^*\mu^*$

An approximated numerical system

Every numerical approximation $\widehat{\varphi}$ satisfying Assumptions 1-2 will give rise to a transition probability $\widehat{P}$ on $(S, \mathbb{S})$. But even if $\widehat{\varphi}$ is an arbitrarily good approximation of function φ, the asymptotic dynamics under transition functions P and $\widehat{P}$ may be quite different.

Biased simulations

Transition function P is defined by the following Markov matrix

$$\Pi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

An element π_{ij} corresponds to the value $P(s_i, \{s_j\})$, for $i,j = 1,2,3$. Note that $\Pi^n = \Pi$ for all $n \geq 1$. Hence, $p = (1,0,0)$, and $p = (0,1/2,1/2)$ are invariant distributions of Π, and $\{s_1\}$ and $\{s_2, s_3\}$ are the ergodic sets.
Now consider an approximation

$$\widehat{\Pi} = \begin{bmatrix} 1-2\delta & \delta & \delta \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix} \text{ for } 0 < \delta < 1/2.$$

Then, as $n \to \infty$ $\{\widehat{\Pi}^n\}$ converges to

$$\begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

Hence, $p = (0,1/2,1/2)$ is the only possible long-run distribution for the system. Moreover, $\{s_1\}$ is a transient state, and $\{s_2, s_3\}$ is the only ergodic set.

Basic Assumptions

- The set S is compact.
- Function $\varphi : S \times E \to S$ is bounded and jointly measurable. Moreover, for every continuous function $f : S \to R$,

$$\int f(\varphi(s_j, \varepsilon))Q(d\varepsilon) \to_j \int f(\varphi(s, \varepsilon))Q(d\varepsilon) \text{ as } s_j \to_j s. \qquad (2.3)$$

An invariant distribution always exists

Theorem

Under Assumptions 1-2, there exists a probability measure μ^* such that $\mu^* = T^*\mu^*$.

To study approximation of moments allow for multiplicity

Let

$$E^{max}(f) = \max_{\{\mu^* | \mu^* = T^* \mu^*\}} \int f(s)\mu^*(ds) \tag{3.3a}$$

$$E^{min}(f) = \min_{\{\mu^* | \mu^* = T^* \mu^*\}} \int f(s)\mu^*(ds). \tag{3.3b}$$

Remark: The set of invariant distributions $\{\mu^* | \mu^* = T^* \mu^*\}$ is weakly compact and convex, we get
$[E^{min}(f), E^{max}(f)] = \{\int f(s)\mu^*(ds) | \mu^* = T^* \mu^*\}$.

Notation

Let $\|\cdot\|$ be the max norm in R^l. Then, for any two vector-valued functions φ and $\widehat{\varphi}$ let

$$d(\varphi, \widehat{\varphi}) = \max_{s \in S}[\int \|\varphi(s, \varepsilon) - \widehat{\varphi}(s, \varepsilon)\| \, Q(d\varepsilon)]. \qquad (3.1)$$

Using this norm, consider a sequence of functions $\{\varphi_j\}$ converging to φ. Note that by Assumptions 1-2 each φ_j defines the associated pair (P_j, T_j^*); and that an invariant distribution exists.

Accuracy in moments

Theorem

Let f belong to $C(S)$. Then, for every $\eta > 0$ there exists J such that

$$E^{min}(f) - \eta < \int f(s)\mu_j^*(ds) < E^{max}(f) + \eta \qquad (3.4)$$

for all μ_j^* with $j \geq J$.

Problem

The previous results are very interesting, but how do we compute invariant distributions and their moments?

Accuracy of simulated moments

We can draw sequences $\{\widehat{\varepsilon}_n\}$. A probability measure λ is defined over all sequences $\omega = (\varepsilon_1, \varepsilon_2, ...)$. Once a numerical approximation φ_j is available, we can generate sample paths $\{s_{jn}(s_0, \omega)\}$ defined recursively as $s_{jn+1}(s_0, \omega) = \varphi_j(s_{jn}(s_0, \omega), \varepsilon_{n+1})$ for every $n \geq 0$ for fixed s_0 and ω. Then we get sequences of simulated statistics $\{\frac{1}{N}\sum_{n=1}^{N} f(s_{jn}(s_0, \omega))\}$, for some function f. Aim: for a sufficiently good numerical approximation φ_j and for a sufficiently large N the series $\{\frac{1}{N}\sum_{n=1}^{N} f(s_{jn}(s_0, \omega))\}$ is close (almost surely) to the expected value $E(f) = \int f(s)\mu^*(ds)$ of some invariant distribution μ^* of the original equilibrium function φ.

Accuracy of simulated moments

Theorem

Under our prevailing assumptions, for every $\eta > 0$ there are functions $N_j(w)$ and an integer J such that for all $j \geq J$ and $N \geq N_j(\omega)$,

$$E^{min}(f) - \eta < \frac{1}{N} \sum_{n=1}^{N} f(s_{jn}(s_0, \omega)) < E^{max}(f) + \eta \qquad (3.8)$$

for all s_0 and λ-almost all ω.

Accuracy of simulated moments

Corollary

Assume that there exists a unique invariant distribution $\mu^* = T^*\mu^*$. *Then for all* $j \geq J$ *and* $N \geq N_j(\omega)$,

$$|\frac{1}{N}\sum_{n=1}^{N} f(s_{jn}(s_0, \omega)) - E(f)| < \eta \qquad (3.10)$$

for all s_0 *and* λ*-almost all* ω.

Observe that each approximating function φ_j may contain multiple invariant distributions μ_j^*.

Contractive systems

CONDITION C: There exists a constant $0 < \gamma < 1$ such that $\int \|\varphi(s,\varepsilon) - \varphi(s',\varepsilon)\| \, Q(d\varepsilon) \leq \gamma \|s - s'\|$ for all pairs s, s'.

Theorem

The true solution of the model has a unique invariant distribution.

Error bounds

Theorem

Let f be a Lipschitz function with constant L. Let $d(\varphi, \widehat{\varphi}) \leq \delta$ for some $\delta > 0$. Assume that φ satisfies Condition C. Then for every $\eta > 0$ there exists a function $\widehat{N}(\omega)$ such that for all $N \geq \widehat{N}(\omega)$,

$$\left|\frac{1}{N}\sum_{n=1}^{N} f(\widehat{s}_n(s_0, \omega)) - \int f(s)\mu^*(ds)\right| \leq \frac{L\delta}{1-\gamma} + \eta \qquad (4.3)$$

for all s_0 and λ-almost all ω.