Neural Network Learning for Nonlinear Economies

Julian Ashwin[1] Paul Beaudry[2] Martin Ellison[3]

[1]Maastricht University

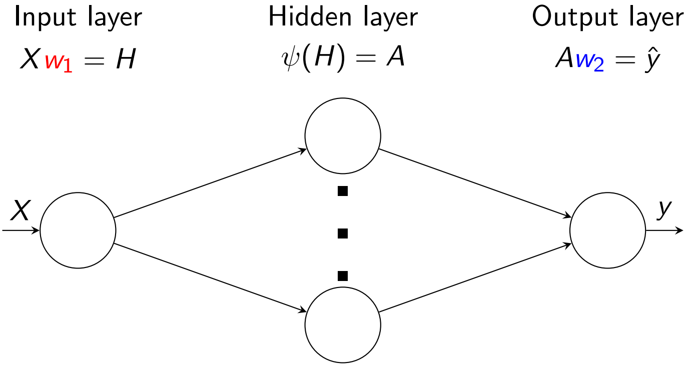
[2]University of British Columbia

[3]University of Oxford and CEPR

Introduction

- Nonlinearities abound in macroeconomics
  - flattening of Phillips curve at low inflation
  - zero lower bound on nominal interest rate
- Any nonlinear REE can be represented by a neural network
  - Universal Approximation Theorem
  - Cybenko 1989, Hornik et al. 1989, Hornik 1991
- We ask if agents can learn the neural network representation
  - application of stochastic approximation theory
  - extends e-stability analysis to nonlinear models

Simple feed-forward neural network



Neural network learning

- Agents use the neural network to form expectations
- Learn parameters w_1,w_2 by method of gradient descent
- Update estimates in response to prediction error $\zeta_t = y_t - \hat{y}_t$

$$
\begin{aligned}
w_1' &= w_1 - \gamma_t \nabla_{w_1} \zeta_t \\
w_2' &= w_2 - \gamma_t \nabla_{w_2} \zeta_t
\end{aligned}
$$

- ∇_{w_1} and ∇_{w_2} are partial derivatives of prediction error
- Learning rate parameter γ_t regulates magnitude of updating

Stochastic approximation

- Updating equations are ordinary difference equation

$$\underbrace{\begin{bmatrix} w_1' \\ w_2' \end{bmatrix}}_{\theta_t} = \underbrace{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{\theta_{t-1}} + \gamma_t \underbrace{\begin{bmatrix} X_t^T w_2^T \psi'(X_t w_1) \\ \psi(X_t w_1)^T \end{bmatrix} \zeta_t}_{Q(\theta_{t-1}, Z_t)}$$

- This is a stochastic recursive algorithm of standard form
- Stochastic approximation theory says we can learn a lot about its properties by associating an ordinary differential equation (ODE)

Associated ODE

$$\frac{d\theta}{d\tau} = h(\theta(\tau))$$
$$h(\theta) = \lim_{t\to\infty} EQ\left(\theta, \bar{Z}_t(\theta)\right)$$

- $h(\theta(\tau))$ is expected update as a function of $\bar{Z}_t(\theta)$
- $\bar{Z}_t(\theta)$ is stochastic process for Z_t holding θ_{t-1} fixed at θ
- Limit point $h(\theta^*) = 0$ represents an equilibrium of system
- Equilibrium learnable if ODE asymptotically stable
- Asymptotic stability requires all eigenvalues of Jacobian $Dh(\theta)$ to have negative real parts when evaluated at θ^*

A target range for inflation

- FOMC discussed adoption of target range for inflation in September 2019 when deliberating on future of US monetary policy framework
- Target range "the hallmark of inflation targeting" (Bernanke, 2003)
- Recent research by Bianchi et al. 2021, Le Bihan et al. 2023
- Implicit assumption is that monetary policy reacts differently depending on whether inflation is inside or outside target range

Analytic example

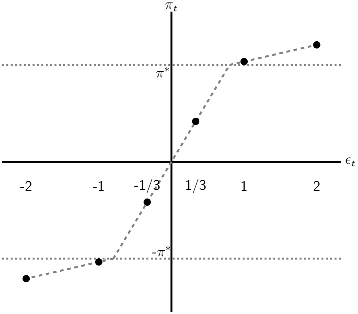
$$
\begin{aligned}
\pi_t &= \mathsf{E}_t \pi_{t+1} + \kappa y_t + \epsilon_t \\
y_t &= -\sigma(R_t - \mathsf{E}_t \pi_{t+1}) \\
R_t &= \begin{cases} \alpha(\pi_t + \pi^*) & \text{if } \pi_t < -\pi^* \\ 0 & \text{if } -\pi^* \leq \pi_t \leq \pi^* \\ \alpha(\pi_t - \pi^*) & \text{if } \pi_t > \pi^* \end{cases}
\end{aligned}
$$

- Disturbance term ϵ_t follows 6-state Markov chain
- Probability of remaining in same state is p
- Probability of switching to each other state is $(1 - p)/5$

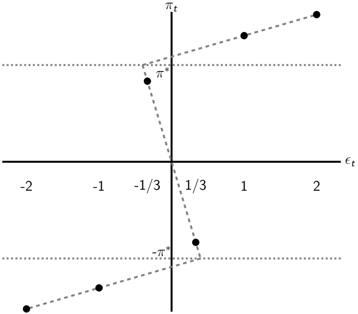
REE

- One state variable so REE characterised by π_t at each ϵ_t (abstracting for now from sunspot solutions)
- Start with REE in which inflation sometimes outside target range
- Guess and verify to obtain unique REE
- Nature of unique REE depends on persistence p of ϵ_t
  - Monotonic if $p < p^*$
  - Non-monotonic if $p > p^*$
- Highly persistent ϵ_t disinflationary in target range
  - $R_t = 0$ and $E_t\pi_{t+1} \approx \pi_t \rightarrow y_t \approx \sigma\pi_t$ and $\pi_t \approx -\epsilon_t/(\sigma\kappa)$

REE with inflation sometimes outside the target range



(a) ϵ_t mildly persistent



(b) ϵ_t highly persistent

Learning

- REE piecewise linear with two breakpoints
- Sufficient to have neural network with one node always activated and two nodes activated by ReLU functions
- Neural network representation

$$E_t \pi_{t+1} = b_{21}(a_1 + \epsilon_t) + b_{22}\max(a_2 + \epsilon_t, 0) + b_{23}\max(a_3 + \epsilon_t, 0)$$

- Supports REE for $w_1^* \equiv (a_1^*, a_2^*, a_3^*)$ and $w_2^* \equiv (b_{21}^*, b_{22}^*, b_{23}^*)$
- Six parameters to learn by method of gradient descent

Associated ODE

- Stochastic recursive algorithm of standard form

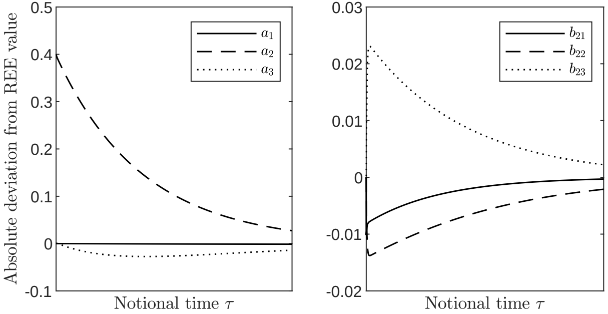
$$\underbrace{\begin{bmatrix} w_1' \\ w_2' \end{bmatrix}}_{\theta_t} = \underbrace{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}}_{\theta_{t-1}} + \gamma_t \underbrace{\begin{bmatrix} X_t^T w_2^T \psi'(X_t w_1) \\ \psi(X_t w_1)^T \end{bmatrix} \zeta_t,}_{Q(\theta_{t-1}, Z_t)} \tag{1}$$

- Prediction errors $\zeta_t = \pi_t - E_{t-1}\pi_t$ depend on ϵ_t and ϵ_{t-1}
- Six states for ϵ_t so $6^2 = 36$ possible prediction errors
- Six parameters so $h(\theta) = \lim_{t\to\infty} EQ\left(\theta, \bar{Z}_t(\theta)\right)$ is 6×1 vector
- Jacobian $Dh(\theta)$ is 6×6 matrix

Learnability

- $h(\theta^*) = 0$ at REE parameter values w_1^* and w_2^*
- Learnability requires all eigenvalues of Jacobian $Dh(\theta)$ to have negative real parts when evaluated at θ^*
- Six eigenvalues so require $|Dh(\theta)| > 0$ at θ^*
- Necessary condition $p < p^*$
- REE not learnable if persistence of ϵ_t is high
- REE could be learnable if persistence of ϵ_t is mild

Learnability of REE with mild persistence



Non-learnability of REE with high persistence

- Neural network representation

$$\mathrm{E}_t \pi_{t+1} = b_{21}(a_1 + \epsilon_t) + b_{22}\max(a_2 + \epsilon_t, 0) + b_{23}\max(a_3 + \epsilon_t, 0)$$

- Consider perturbation to a_2 that increases inflation expectations when inflation is in its target range
- Increase in inflation expectations raises inflation more than proportionately as long as inflation stays in the target range
- Prediction error $\zeta_t = \pi_t - E_{t-1}\pi_t > 0$
- Update to a_2 takes it even further away from its REE value

Aside on REE with sunspot fluctuations

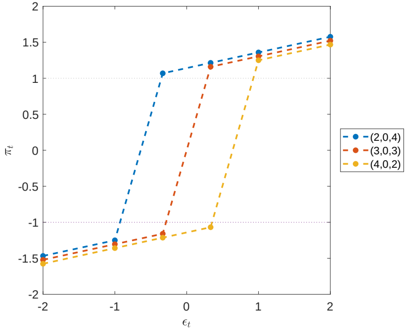
- REE with sunspot fluctuations also exist
- Sunspot η_t attached to $E_t\pi_{t+1}$ when inflation in target range
- η_t switches between $\pm\eta$ with 'resonant frequency' $q(p, \sigma, \kappa)$
- REE has neural network representation

$$
\begin{aligned}
\mathrm{E}_t\pi_{t+1} = & b_{21}(a_1 + b_{11}\epsilon_t + c_1\eta_t) + b_{22}\max(a_2 + b_{12}\epsilon_t, 0) \\
& + b_{23}\max(a_3 + b_{13}\epsilon_t, 0),
\end{aligned}
$$

- $h(\theta)$ now a 7×1 vector and $Dh(\theta)$ a 7×7 matrix
- $p > p^*$ when 'resonant frequency' condition satisfied
- Condition that facilitates sunspot fluctuations in REE precludes agents and neural network from learning them

REE with inflation always outside the target range

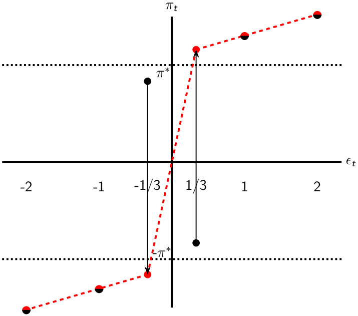
- REE exist where inflation always outside target range
- Guess and verify to obtain REE



- Each has neural network representation

Learnability

- $h(\theta)$ is 4×1 vector and $Dh(\theta)$ is 4×4 matrix
- $|Dh(\theta)| > 0$ and Descartes' rule of signs $\rightarrow$ no positive roots at θ^*
- REE with inflation always outside target range are learnable



Numerical example with target range for inflation

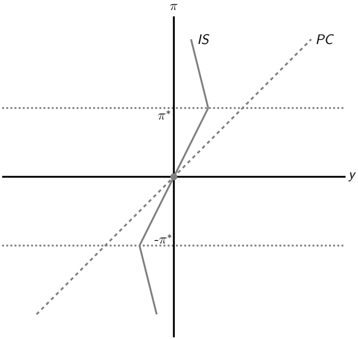
$$
\pi_t = \beta \mathsf{E}_t \pi_{t+1} + \kappa y_t + \epsilon_{\pi,t}
$$

$$
y_t = \eta y_{t-1} - \sigma(R_t - \mathsf{E}_t \pi_{t+1}) + \epsilon_{y,t}
$$

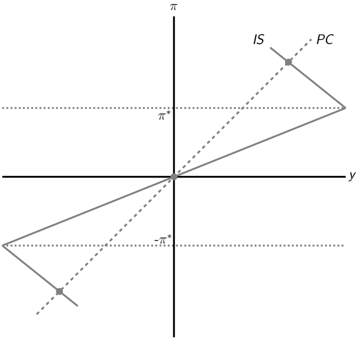
$$
R_t = \begin{cases} \phi_\pi \pi_t + \alpha(\pi_t + \pi^*) & \text{if } \pi_t < -\pi^* \\ \phi_\pi \pi_t & \text{if } -\pi^* \leq \pi_t \leq \pi^* \\ \phi_\pi \pi_t + \alpha(\pi_t - \pi^*) & \text{if } \pi_t > \pi^* \end{cases}
$$

β	κ	η	σ	ϕ_π	α	ρ_π	ρ_y	σ_π	σ_y
0.95	0.05	(0.75,0.95)	0.25	0.5	0.75	0.5	0.5	0.2	0.2

Deterministic steady states

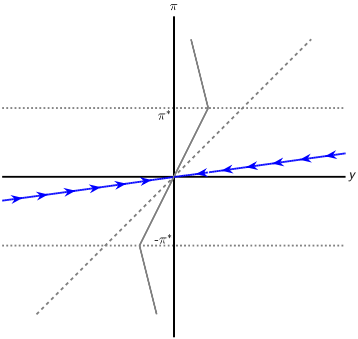


(a) η low



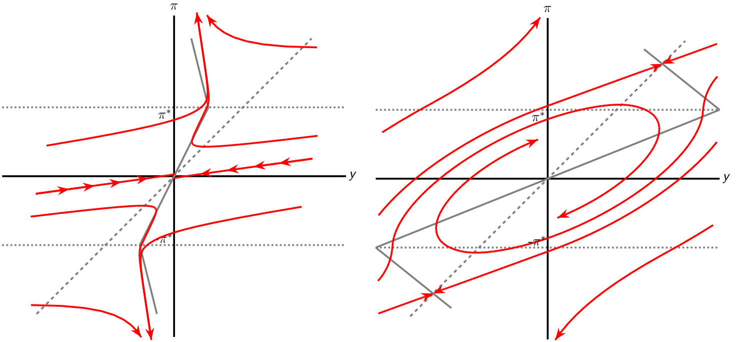
(b) η high

Equilibrium dynamics under neural network learning



(a) η low (b) η high

Perfect foresight



(a) η low (b) η high

Figure: Perfect foresight paths

Empirical tests for indeterminacy

- Apply Lubik and Schorfheide (2004) test to simulated data from model under neural network learning
- Compare probability data generated by a locally-determinate system as opposed to perturbations of perfect foresight paths in a locally-indeterminate system
- Learnable REE dynamics resemble neither saddlepath stable system nor perturbations to perfect foresight paths when inertia is high
- Can test misleadingly favour sunspots as a driver of simulated data?

Recap on Lubik and Schorfheide (2004)

- Solve log-linearised model under RE

$$
\begin{aligned}
\mathbb{E}_t \tilde{x}_{1,t+1} &= \Lambda_1 \mathbb{E}_{t-1} \tilde{x}_{1,t} + \Lambda_1 (\tilde{x}_{1,t} - \mathbb{E}_{t-1} \tilde{x}_{1,t}) + \tilde{\epsilon}_{1,t}, \\
\mathbb{E}_t \tilde{x}_{2,t+1} &= \Lambda_2 \mathbb{E}_{t-1} \tilde{x}_{2,t} + \Lambda_2 (\tilde{x}_{2,t} - \mathbb{E}_{t-1} \tilde{x}_{2,t}) + \tilde{\epsilon}_{2,t}.
\end{aligned}
$$

- Indeterminacy if Λ_1 or Λ_1 both inside unit circle
- Sunspot process ζ_t has to respect RE s.t. $\zeta_t = \pi_t - E_{t-1}\pi_t$
- ζ_t can be correlated with innovations to disturbance terms

$$
\mathbb{E}_{t-1} \begin{pmatrix} \epsilon_{\pi,t} \\ \epsilon_{y,t} \\ \zeta_t \end{pmatrix} \begin{pmatrix} \epsilon_{\pi,t} \\ \epsilon_{y,t} \\ \zeta_t \end{pmatrix}' = \begin{pmatrix} \sigma_\pi & 0 & \omega_{\pi,\zeta} \\ 0 & \sigma_y & \omega_{y,\zeta} \\ \omega_{\pi,\zeta} & \omega_{y,\zeta} & \sigma_\zeta \end{pmatrix}
$$

Results

Parameter	Prior mean	Posterior mean for $\eta = 0.75$		Posterior mean for $\eta = 0.95$	
		Determinate	Indeterminate	Determinate	Indeterminate
β	$\mathcal{N}(1,1)$	0.06 (−0.57,0.84)	1.90 (1.77,2.00)	0.40 (0.26,0.47)	1.37 (1.20,1.57)
κ	$\mathcal{N}(0,1)$	0.16 0.07,0.25)	−0.10 (−0.14−0.05)	0.48 (0.35,0.58)	−0.26 (−0.42,0.14)
η	$\mathcal{N}(1,1)$	0.75 (0.73,0.78)	0.75 (0.73,0.77)	0.98 (0.97,0.99)	0.60 (0.49,0.70)
σ	$\mathcal{N}(0,1)$	−0.09 (−0.22,0.00)	0.15 (0.01,0.36)	0.001 (−0.01,0.02)	0.65 (0.52,0.80)
ϕ_π	$\mathcal{N}(2,1)$	1.32 (0.62,1.94)	0.68 (0.38,0.99)	1.81 (1.64,1.98)	0.22 (0.08,0.39)
ρ_π	$\mathcal{N}(0,1)$	0.51 (0.49,0.52)	0.24 (−0.05,0.54)	0.50 (0.48,0.52)	0.42 (0.38,0.47)
ρ_y	$\mathcal{N}(0,1)$	0.49 (0.45,0.52)	0.50 (0.47,0.53)	0.48 (0.47,0.49)	0.51 (0.48,0.55)
σ_π	$\mathcal{IG}(0.5,2)$	0.37 (0.22,0.49)	0.31 (0.17,0.46)	0.31 (0.28,0.34)	0.24 (0.21,0.27)
σ_y	$\mathcal{IG}(0.5,2)$	0.21 (0.20,0.21)	0.21 (0.20,0.21)	0.21 (0.21,0.21)	0.14 (0.12,0.15)
σ_ζ	$\mathcal{IG}(0.5,2)$	-	0.34 (0.31,0.38)	-	0.41 (0.40,0.43)
$\omega_{\pi,\zeta}$	$\mathcal{B}(0,0.3,-1,1)$	-	. (−,+)	-	0.39 (0.19,0.58)
$\omega_{y,\zeta}$	$\mathcal{B}(0,0.3,-1,1)$	-	. (−,+)	-	−0.10 (−0.46,0.28)
Log data density		**-3012**	−3019	−3499	**-3496**

Numerical example with a Zero Lower Bound

- Evans et al. (2023)
- Add consumption habits and endogenous government spending
- Continuum of identical household-producers indexed by i

$$\mathrm{E}_{0,i} \sum_{t=0}^{\infty} \beta^t \{ \log(c_{t,i} + \xi_t g_t - \lambda c_{t-1}) + \chi \log \left(\frac{M_{t-1,i}}{P_t} \right) \\ - (1+\epsilon)^{-1} h_{t,i}^{1+\epsilon} - \Phi \left(\frac{P_{t,i}}{P_{t-1,i}} \right) \}$$

- $\Phi(\cdot) \geq 0$ is a convex pricing friction

Numerical example with a Zero Lower Bound

- Nominal interest rate

$$R_t = \max\left\{ R^* \left(\frac{\pi_t}{\pi^*}\right)^{\phi_\pi} \left(\frac{y_t}{y^*}\right)^{\phi_y}, 1 \right\},$$

- Government spending

$$g_t = \frac{\bar{g}}{1 + e^{k(\pi_t - \pi^*)}}.$$

- Lump-sum taxes balance period budget constraint

$$b_t + m_t + \Upsilon_t = g_t + (m_{t-1} + R_{t-1} b_{t-1})/\pi_t$$

Equilibrium conditions

- Euler equation for consumption

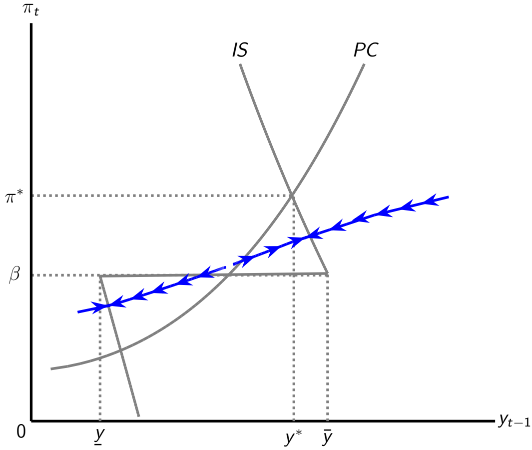
$$c_t = \max(c_t^*, 0)$$

$$\frac{1}{c_t^* + \xi_t g_t - \lambda c_{t-1}} = \beta R_t \mathrm{E}_t \left(\frac{\pi_{t+1}^{-1}}{c_{t+1}^* + \xi_{t+1} g_{t+1} - \lambda c_t} \right)$$

- Nonlinear New Keynesian Phillips curve

$$\Phi'(\pi_t)\pi_t = \frac{\nu}{\alpha} \left(\frac{c_t + g_t}{A_t} \right)^{\frac{\epsilon+1}{\alpha}} + \frac{1-\nu}{c_t + \xi_t g_t - \lambda c_{t-1}} (c_t + g_t) + \beta \mathrm{E}_t \Phi'(\pi_{t+1})\pi_{t+1}$$

Deterministic steady states and equilibrium dynamics



Conclusions

- Neural network learning is a practical and powerful tool
- Any REE has a neural network representation
- Learnability of any REE can be checked analytically or numerically
- A solution technique and equilibrium selection device
- High persistence drives inflation away from target range
- REE with sunspot fluctuations typically not learnable
- Inflation pushed outwards if there is a ZLB