Discussion of “Behavioral Macroeconomics Via Sparse Dynamic Programming” by Xavier Gabaix

Jennifer La’O
Columbia and NBER

July 16, 2016

Sparse Dynamic Programming

- Agent builds a "sparse" simplified model of the world
    - pays attention only to the most important variables
    - pays zero attention to less important variables
- This Paper: Application to Dynamic Programming and Macro

This Discussion

- General Sparsity Framework: Static Example
- Dynamic Two-Period Application
- Infinite Horizon, Dynamic Programming Problem
- My Comments

Sparsity General Framework:

Simple Static Example

Single Agent Static Maximization Problem

$$\max_a v\left(a, \mathbf{x}\right)$$

- suppose a is a single action
- $\mathbf{x} = \left(x_1, x_2, \ldots, x_n\right)$ is a vector of state variables

$$x_i \sim \mathcal{N}(0, \sigma_i^2)$$

- but perhaps the set of x_i is really, really, really big!

Alternative: Sparsity (Gabaix 2014 QJE)

- agent instead chooses a *sparse* representation of x_i

$$x_i^s = m_i x_i$$

with attention vector

$$\mathbf{m} = (m_1, m_2, \ldots, m_n) \quad \text{where} \quad m_i \in [0, 1]$$

  - if $m_i < 1$ then agent pays only partial attention to x_i
  - if $m_i = 0$ then agent pays no attention to x_i

- agent maximizes new, sparse objective

$$v(a, \mathbf{x}) \to v(a, m_1 x_1, \ldots, m_n x_n)$$

Step 1: Choose attention vector

- minimize utility loss due to inattentiveness

$$\min_{m_i} \frac{1}{2}\Lambda_i (1 - m_i)^2 + \kappa m_i$$

- Λ_i denotes expected utility loss

  (and here I assume x_i are uncorrelated)

- κ constant cost of attention

Utility Loss evaluated at Default Model

- How do you calculate Λ_i?

$$\Lambda_i \equiv -v_{aa} \left(\frac{\partial a}{\partial x_i} \right)^2 \sigma_i^2$$

- derivatives evaluated around a "Default Model"

$$v_{aa} = \left. \frac{\partial^2 v}{\partial a^2} \right|_{a=a^d} \quad \text{and} \quad \left. \frac{\partial a}{\partial x_i} \right|_{a=a^d}$$

- Default Model: can choose, but for most applications, $m_i^d = 0$

Optimal Attention Choice

$$\min_{m_i} \frac{1}{2}\Lambda_i (1 - m_i)^2 + \kappa m_i$$

- FOC

$$\kappa = \Lambda_i (1 - m_i)$$

- solve for m to obtain optimal attention coefficient

$$m_i^* = \max \left\{ 1 - \frac{\kappa}{\Lambda_i}, 0 \right\}$$

Optimal Attention is Sparse

$$m_i^* = \begin{cases} 0 & \text{if } \kappa > |v_{aa}| \left(\frac{\partial a}{\partial x_i}\right)^2 \sigma_i^2 \\ 1 - \frac{\kappa}{|v_{aa}| \left(\frac{\partial a}{\partial x_i}\right)^2 \sigma_i^2} & \text{otherwise} \end{cases}$$

"Eliminate attention if the marginal cost of attention is too high relative to the marginal benefit"

Step 2: Solve problem with sparse representation

$$\max_{a} v\left(a, m_1^* x_1, \ldots, m_n^* x_n\right)$$

Sparsity in a Two-Period Model

Single Agent Consumption-Savings Problem

$$\max_{c_0, c_1} u(c_0) + \mathbb{E}\beta u(c_1)$$

subject to

$$\begin{aligned} c_0 + a_1 &\leq y_0 \\ c_1 &\leq (1 + r) a_1 + y_1 \end{aligned}$$

where

$$\begin{aligned} r &= \bar{r} + \varepsilon_r & \varepsilon_r \sim \mathcal{N}(0, \sigma_r^2) \\ y_1 &= \bar{y}_1 + \varepsilon_y & \varepsilon_y \sim \mathcal{N}(0, \sigma_y^2) \end{aligned}$$

Sparse Model

$$v\left(a_1, m_r\varepsilon_r, m_y\varepsilon_y\right) = u\left(y_0 - a_1\right) + \beta u\left(\left(1+r\right)a_1 + y_1\right)$$

where

$$\begin{aligned} r &= \bar{r} + m_r\varepsilon_r \\ y_1 &= \bar{y}_1 + m_y\varepsilon_y \end{aligned}$$

Default Model

- default model:

$$m_r = 0, \quad m_y = 0$$

so that agent completely ignores the stochastic components

$$r = \bar{r} \quad \text{and} \quad y_1 = \bar{y}_1$$

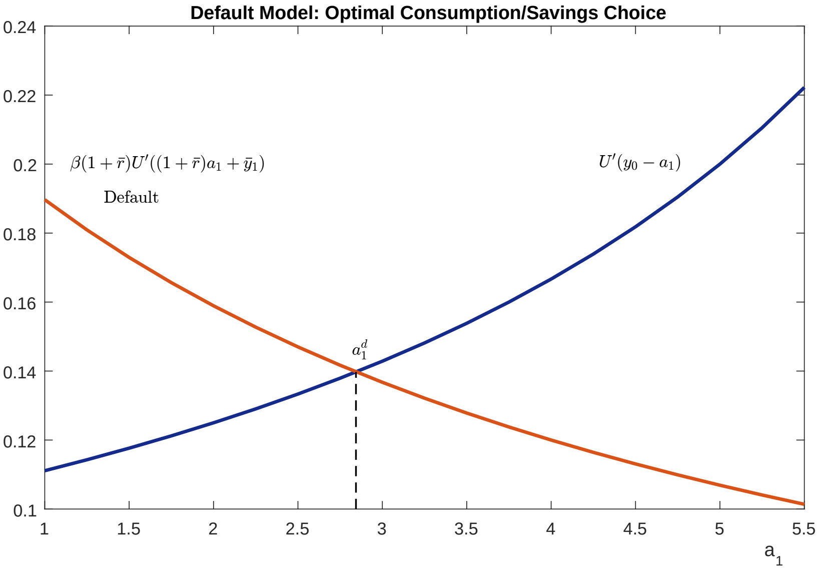
- optimality condition: Euler Equation

$$u'(c_0) = \beta (1 + \bar{r}) u'(c_1)$$

- one line solution

$$u'(y_0 - a_1) = \beta (1 + \bar{r}) u'((1 + \bar{r}) a_1 + \bar{y}_1)$$

Default Model Solution



Rational Agent

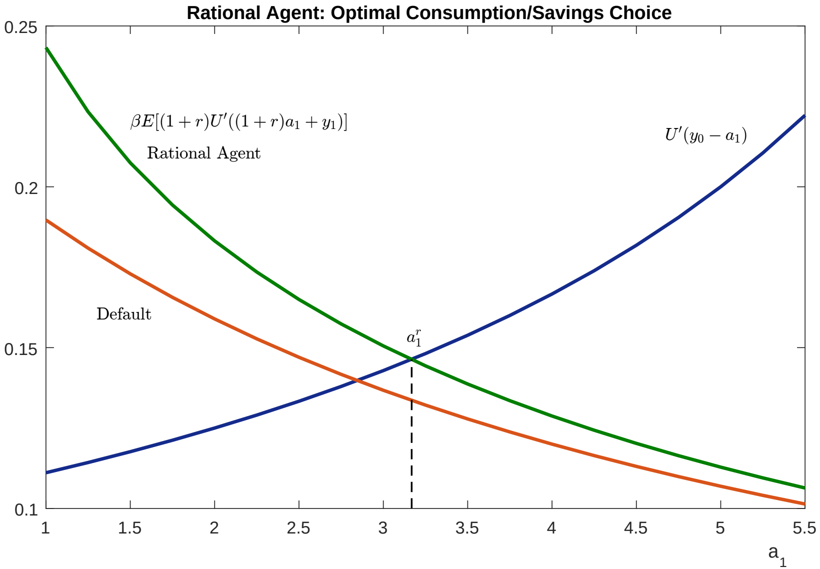
- Euler equation

$$u'(c_0) = \beta \mathbb{E}(1+r)\,u'(c_1)$$

- one line solution

$$u'(y_0 - a_1) = \beta \mathbb{E}\left[(1+r)\,u'((1+r)\,a_1 + y_1)\right]$$

Rational Agent Solution



Sparse Agent. Step 1: Optimal Attention

- calculate Λ_i expected utility loss

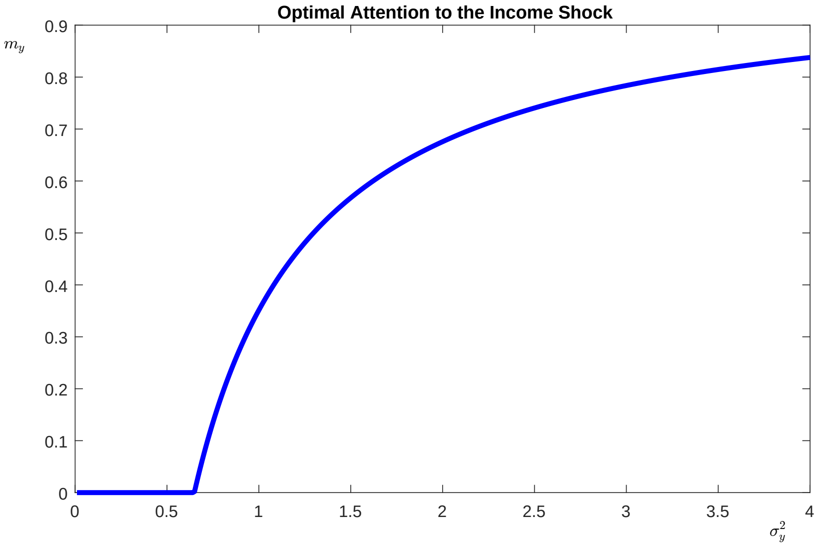
$$\Lambda_i = -v_{aa} \left(\frac{\partial a}{\partial x_i} \right)^2 \sigma_i^2$$

  with derivatives evaluated around the Default Model

- obtain optimal attention coefficient

$$m_i^* = \max \left\{ 1 - \frac{\kappa}{\Lambda_i}, 0 \right\}$$

Optimal Attention as a function of Variance



Optimal Attention as a function of Cost

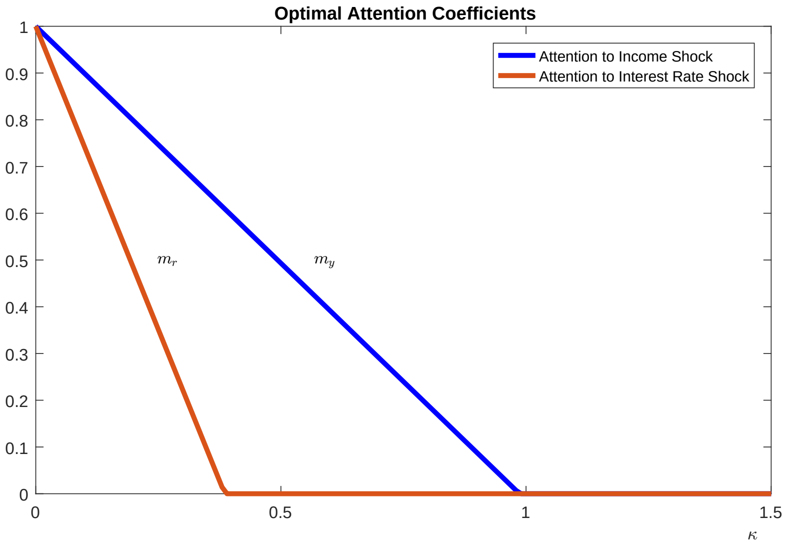


Figure:

Sparse Agent. Step 2: Consumption-Savings Choice.

- given m_r^* and m_y^*, now solve

$$\max_{a_1} v\left(a_1, m_r^* \varepsilon_r, m_y^* \varepsilon_y\right)$$

- that is,

$$\max_{a_1} u\left(y_0 - a_1\right) + \beta \mathbb{E} u\left(\left(1+r\right) a_1 + y_1\right)$$

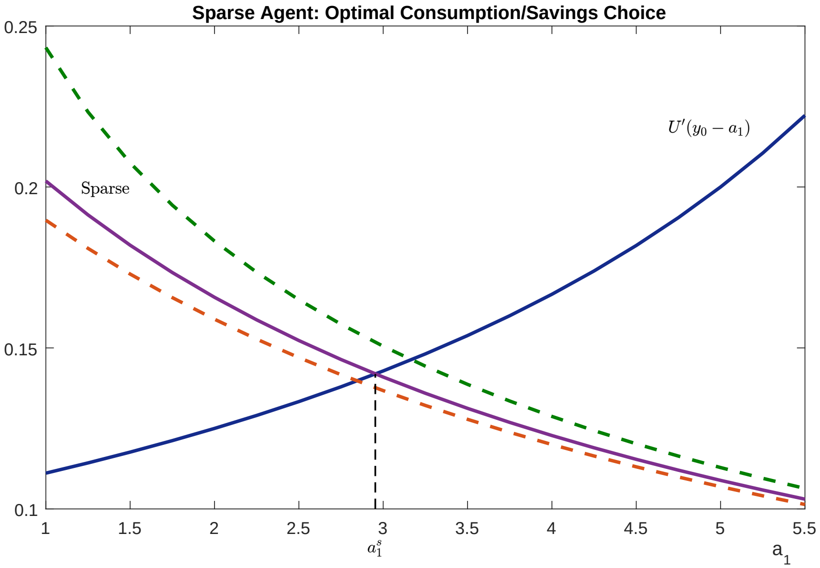
where

$$\begin{aligned} r &= \bar{r} + m_r^* \varepsilon_r \\ y_1 &= \bar{y}_1 + m_y^* \varepsilon_y \end{aligned}$$

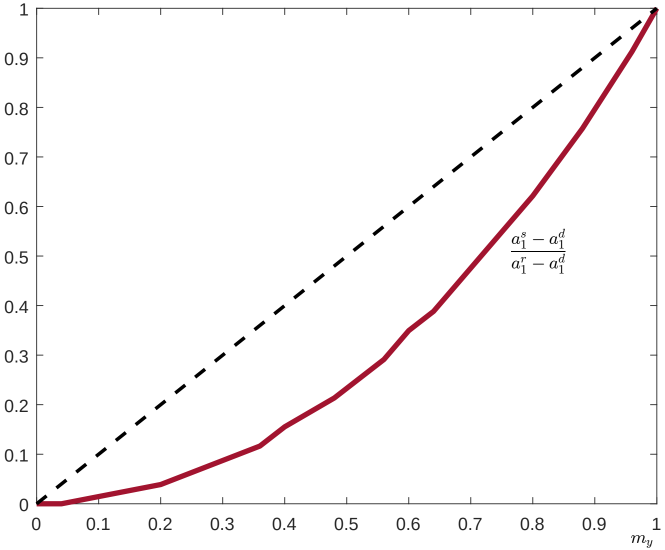
- one line solution

$$u'\left(y_0 - a_1\right) = \beta \mathbb{E}\left[\left(1+r\right) u'\left(\left(1+r\right) a_1 + y_1\right)\right]$$

Sparse Agent Solution



Action isn't linear in attention



Implications

- for intermediate values of attention cost,
  - agent pays no attention to interest rate shock
  - pays less than full attention to income shock
- optimal action, savings, is attenuated relative to fully rational model
  - agent builds less precautionary savings
- affects estimation/measurement of IES

Infinite-Horizon Model with

Sparse Dynamic Programming

Infinite Horizon Model

$$\max_{\{c_t\}} \mathbb{E} \sum_{t=0}^{\infty} \beta^t u(c_t)$$

subject to

$$w_{t+1} \leq (1 + \bar{r} + \hat{r}_{t+1})(w_t - c_t) + \bar{y} + \hat{y}_{t+1}$$

where

$$\begin{aligned}
\hat{r}_{t+1} &= \rho_r \hat{r}_t + \varepsilon^r_{t+1}, \qquad \varepsilon^r_{t+1} \sim \mathcal{N}(0, \sigma_r^2) \\
\hat{y}_{t+1} &= \rho_y \hat{y}_t + \varepsilon^y_{t+1}, \qquad \varepsilon^y_{t+1} \sim \mathcal{N}(0, \sigma_y^2)
\end{aligned}$$

Rational Model

- state space

$$\left(w_t, \hat{r}_t, \hat{y}_t\right)$$

- bellman equation

$$V\left(w, \hat{r}, \hat{y}\right) = \max_a \left\{u\left(a, w, \hat{r}, \hat{y}\right) + \beta E V\left(w', \hat{r}', \hat{y}'\right)\right\}$$

Sparse Model

- Gabaix proposes the following law of motion for states

$$w_{t+1} = \left(1 + \bar{r} + m_r \hat{r}\right)\left(w_t - c_t\right) + \bar{y} + m_y \hat{y}_t$$

and

$$\begin{aligned} \hat{r}_{t+1} &= \rho_r(m)\,\hat{r}_t + m_{\sigma r}\varepsilon^r_{t+1}, \quad \text{where} \quad \rho_r(m) = m_{\rho r}\rho_r + \left(1 - m_{\rho r}\right)\rho^d_r \\ \hat{y}_{t+1} &= \rho_y(m)\,\hat{y}_t + m_{\sigma y}\varepsilon^y_{t+1}, \quad \text{where} \quad \rho_y(m) = m_{\rho y}\rho_y + \left(1 - m_{\rho y}\right)\rho^d_y \end{aligned}$$

where $\rho^d_r, \rho^d_y \in [0, 1]$ are some default persistence parameters

- six attention coefficients

$$\mathbf{m} = \left(m_r, m_{\rho r}, m_{\sigma r}; m_y, m_{\rho y}, m_{\sigma y}\right)$$

Comments

Comment 1: Disciplining Sparse Dynamic Programming

- Given static sparsity framework, in dynamic model:
  - what is the default model?
  - where can you place m?
- Natural to choose default model = true non-stochastic steady state

$$\rho_r^d = \rho_r, \qquad \hat{r}_{t+1} = \rho_r \hat{r}_t + m_{\sigma r} \varepsilon_{t+1}^r$$

  - two attention coefficients:

$$m_{\sigma r} \quad \text{and} \quad m_{\sigma y}$$

Disciplining Sparse Dynamic Programming

- With proposed formulation,

$$\mathbf{m} = \left(m_r, m_{\rho r}, m_{\sigma r}; m_y, m_{\rho y}, m_{\sigma y}\right)$$

if $\rho_r^d = 0$ then

$$\rho_r\left(m\right) = m_{\rho r}\rho_r$$

- this introduces cognitive discounting of future variables

$$E\left[\hat{r}_{t+\tau}\right] = m_{\rho r}^{\tau}\rho_r^{\tau}\hat{r}_{t+\tau} < \rho_r^{\tau}\hat{r}_{t+j}$$

- departure from "information processing" interpretation
- Even if model is completely deterministic, agents have the wrong perception of future variables

Departure from Rational Life-Cycle Model

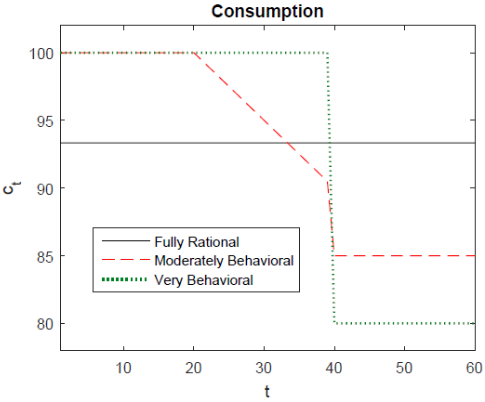


Figure: Sparsity in Life-Cycle

Sparse Bellman Equation

- how do you formulate the value function?
- in rational model, invariance of value function

$$V\left(w,\hat{r},\hat{y}\right)=V\left(w,\hat{r},\hat{y},\ \text{anything else}\right)$$

- in sparsity model with default=true non-stochastic model,

  same invariance of value function, just different measure for shocks

Sparse Bellman Equation

- how do you formulate the value function?
- in rational model, invariance of value function

$$V(w, \hat{r}, \hat{y}) = V(w, \hat{r}, \hat{y}, \mathbb{E}\hat{r}_{t+\tau}, \mathbb{E}\hat{y}_{t+\tau}, \ldots)$$

- in sparsity model with default=true non-stochastic model,

  same invariance of value function, just different measure for shocks

Sparse Bellman Equation

- in proposed sparsity model, with

$$\mathbf{m} = \left(m_r, m_{\rho r}, m_{\sigma r}; m_y, m_{\rho y}, m_{\sigma y}\right)$$

suppose instead

$$V\left(w, \hat{r}, \hat{y}\right) = V\left(w, \hat{r}, \hat{y}, \mathbb{E}\hat{r}_{t+\tau}, \mathbb{E}\hat{y}_{t+\tau}, \ldots\right)$$

where $\mathbb{E}\hat{r}_{t+\tau}, \mathbb{E}\hat{y}_{t+\tau}$ are the true future means of these processes

- if agent then chooses m's, he could to pay attention to these
- what stops agents from paying attention to certain sufficient statistics?

Parsimony with Dynamic Programing

1. Either impose more discipline from original Sparsity Theory
2. Or, change original theory in a way that allows for these departures

Comment 2: When Sparsity gains Traction

- When would *Sparsity* be the most useful?
- complex network models of interactions between many agents
- large production input-output models
    - actions are linear or log-linear
    - implies linear sufficient statistics $\longrightarrow$ sparsity not needed
- but network models of interconnected banks, households
    - contracts are not linear (debt contracts)
    - set of all individual states cannot be reduced
    - sparsity seems almost generically necessary!

Network of CDS Contracts (Boyarchenko, Costello, La'O, Shachar)

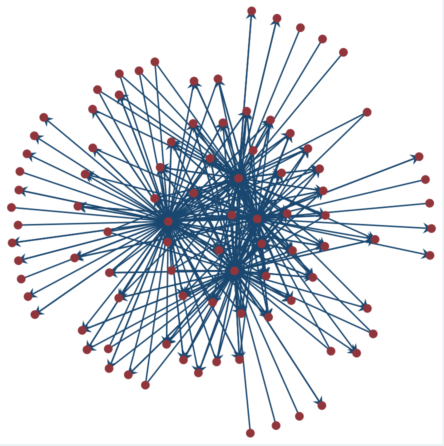


Figure: Inter-Dealer CDS Network (unweighted), January 2015

Last Comment: Technicalities with Feynman-Kac Method

Proof The proof yields the general method of calculation. We shut down uncertainty and differentiate the rational Bellman equation (68), first with respect to the new variable x (using the envelope theorem):

$$V_x(w,x) = u_x + \beta V'_{w'} F^w_x + \beta V'_{x'} F^x_x$$

which yields the announced expression for V_x. Then we take the total derivative w.r.t. w:

$$V_{w,x}(w,x) = D_w u_x + \beta D_w \left[V'_{w'} F^w_x(w,x,a)\right] + \beta F^x_x V'_{w',x'} D_w w' + \beta V'_{x'} D_w F^x_x$$

so

$$V_{w,x}(w,0) = \frac{D_w u_x + \beta D_w \left[F^w_x(w,0,a) V'_{w'}(w',0)\right] + \beta V'_{x'} D_w F^x_x}{1 - \beta F^x_x D_w w'}.$$

□

Lemma 10.2 *Assume the local autonomy condition (32), $F^x_a = 0$, and consider some value function $V(w,x)$. Then, the impact of a change x on the optimal action is (around $x = 0$):*

$$a_x = -\Psi_a^{-1} \Psi_x \tag{72}$$

with $\Psi(a,x) = u_a + \beta V'_{w'} F^w_a$, and

$$\Psi_a = u_{aa} + \beta F^w_a V'_{w'w'} F^w_a + \beta V'_{w'} F^w_{aa}, \qquad \Psi_x = u_{ax} + \beta F^w_x V'_{w'w'} F^w_a + \beta F^x_x V'_{w'x'} F^w_a + \beta V'_{w'} F^w_{ax}. \tag{73}$$

evaluated at $(a,x) = \left(a^d(w), 0\right)$. They depend only on the transition functions and the derivatives of the simpler baseline value function $V^d(w')$.

Proof The FOC for a is $\Psi(a,x) = 0$ with

$$\Psi(a,x) = u_a + \beta V'_{w'} F^{w'}_a + \beta V'_{x'} F^{x'}_a$$

The rest follows by the implicit function theorem: terms in V_{xx} drop out because $F^{x'}_a = 0$ around the default action. □

When the local autonomy condition (32) doesn't hold, a term V_{xx} appears. Then, the situation is more complex, and requires solving for a fixed point, in the form of a matrix Ricatti equation.

Feynman-Kac method In some cases, it is useful to do the same via a Feynman-Kac type of approach.[67] Here we view x_t as exogenous, i.e. assume $x_{t+1} = F^x(x_t)$. Calling w the initial condition for wealth, the Lagrangian is:

$$L = \sum_{t=0}^{\infty} \beta^t u(w_t, x_t, a_t) + \sum_{t=1}^{\infty} \beta^t q^w_t \left(-w_t + F^{w_t}(w_{t-1}, x_{t-1}, a_{t-1})\right) + q^w_0(-w_0 + w)$$

where $q^w_t = V_w(w_t)$ are the Lagrange multipliers. At the optimum, the agent solves $V(w) = \max_{(a_t, w_t)_{t\geq 0}} L$. This implies that $L_{a_t} = L_{w_t} = 0$. The envelope theorem gives:

$$V_{x_t} = L_{x_t} = \beta^t \left[u_x(w_t, x_t, a_t) + \beta q^w_{t+1} F^{w_{t+1}}_x(w_t, x, a_t)\right]$$

so that, using the total derivative notation (69),

$$V_{x_t, w_0} = \beta^t \frac{D w_t}{D w_0}_{|x=0} D_{w_t} \left[u_x\left(w_t, x_t, a^d_t(w_t)\right)_{|x_t=0} + \beta V_w(w_{t+1}) F^{w_{t+1}}_x \left(w_t, x_t, a^d_t(w_t)\right)_{|x_t=0}\right]$$

or in short

$$V_{x_t, w_0} = \sum_{t=0}^{\infty} \beta^t \frac{D w_t}{D w_0} D_{w_t} \left[u^t_x + \beta V^{t+1}_{w_{t+1}} F^{w_{t+1}}_x\right] \tag{74}$$

Application. In the consumption problem with $\beta R = 1$, let us derive again at the impact of a one-time change of interest rate $\hat{r}_t$ from Lemma 4.2. Under the default model $c_t = c_0$ and $w_t = w_0$. So $\frac{D w_t}{D w_0} = 1$, $u_{x_t} = 0$, $V^d_w(w_t) = u'(c_t) = \left(\frac{\bar{r} w_t + \bar{y}}{R}\right)^{-\gamma}$, and given $F^{w_{t+1}} = (1 + r_t)(w_t - c_t) + \bar{y}$,

$$F^{w_{t+1}}_{r_t} = w_t - c_t = w_t - \frac{r w_t + \bar{y}}{R} = \frac{w_t - \bar{y}}{R} = \frac{w^-_t}{R}$$

using the notation $w^-_t := w_t - \bar{y}$ for the beginning of period wealth. Using (74) gives:

$$V_{x_t, w_0} = \beta^t D_{w_t} \left[\beta \left(\frac{\bar{r} w_{t+1} + \bar{y}}{R}\right)^{-\gamma} \frac{w^-_t}{R}\right] = \frac{1}{R^{t+2}} D_{w_t} \left[\left(\frac{\bar{r} w_{t+1} + \bar{y}}{R}\right)^{-\gamma} w^-_t\right]$$
$$= \frac{1}{R^{t+2}} \left(-\gamma \frac{\bar{r}}{R} c_{t+1}^{-\gamma-1} \frac{D w_{t+1}}{D w_t} w^-_t + c_t^{-\gamma}\right) = \frac{1}{R^{t+2}} c_0^{-\gamma-1} \left(-\gamma \frac{\bar{r}}{R} w^-_0 + c_0\right)$$

as under the default model $c_t = c_0$. As time-0 consumption satisfies $u_{c_0} = V_{w_0}$, we have $u_{cc} \partial_{\hat{r}_t} c_0 = \partial_{\hat{r}_t} V_{w_0}$, and

$$\partial_{\hat{r}_t} c_0 = \frac{\partial_{\hat{r}_t} V_{w_0}}{u_{cc}} = \frac{\frac{1}{R^{t+2}} c_0^{-\gamma-1}\left(-\gamma \frac{\bar{r}}{R} w^-_0 + c_0\right)}{-\gamma c_0^{-\gamma-1}} = \frac{1}{R^{t+2}} \left(\frac{\bar{r}}{R} w^-_0 - \psi c_0\right) = \frac{1}{R^{t+2}} \left(\frac{r}{R}(w_0 - \bar{y}) - \psi c_0\right).$$

which gives again Lemma 4.2 (the income part being easy as always).

[67] I call it "Feynman-Kac" because this approach deals particularly well with stochastic problems.

Figure: Appendix for More Complex Model (Section 10) all of you, previous slide:

$$m^*_{\text{this discussion}} = 0$$

Conclusion

Conclusion

- Perfectly Rational Agents (in the strict sense) seems ridiculous
- Sparsity Framework and Agenda: ambitious, creative, also realistic
- Dynamic programming application
    - discipline from General Sparsity Theory
- Most useful in complex network models that are not linear