Dynamic Programming and Structural Estimation

Thomas H. Jørgensen

2023

Outline

1. Introduction
2. Stochastic DP
3. Structural Estimation

Stochastic Dynamic Programming

- **Last time:** Dynamic Programming
  Backwards induction
  Grids
  Interpolation

Stochastic Dynamic Programming

- **Last time:** Dynamic Programming
  Backwards induction
  Grids
  Interpolation
- **Today:**
  - **Uncertainty**:
    Future income is uncertain
    + Another state variable: Permanent income
    + "Normalization" of one state variable.

Stochastic Dynamic Programming

- **Last time:** Dynamic Programming
  Backwards induction
  Grids
  Interpolation
- **Today:**
  - **Uncertainty**:
    Future income is uncertain
    + Another state variable: Permanent income
    + "Normalization" of one state variable.
  - **Estimation:**
    Simulated Method of Moments (SMM/SMD)
    Relate to "reduced-form"
    Can we combine approaches?

Stochastic Dynamic Programming

- **Last time:** Dynamic Programming
  Backwards induction
  Grids
  Interpolation
- **Today:**
  - **Uncertainty**:
    Future income is uncertain
    + Another state variable: Permanent income
    + "Normalization" of one state variable.
  - **Estimation:**
    Simulated Method of Moments (SMM/SMD)
    Relate to "reduced-form"
    Can we combine approaches?
- **Example:** Buffer Stock model of Deaton (1991); Carroll (1992)
  - Estimated in Gourinchas and Parker (2002)

Gourinchas and Parker (2002)

- **Research question:** "Which savings motives dominate across life?"

Gourinchas and Parker (2002)

- **Research question:** "Which savings motives dominate across life?"
- **Approach:**
  1. **Estimate model** with 2+ motives:
     Buffer-stock motive: Income risk while working.
     Life cycle motive: Consumption in retirement.

Gourinchas and Parker (2002)

- **Research question:** “Which savings motives dominate across life?”
- **Approach:**
    1. **Estimate model** with 2+ motives:
       Buffer-stock motive: Income risk while working.
       Life cycle motive: Consumption in retirement.
    2. **Quantify importance** of these motives over life
       Counterfactual simulations

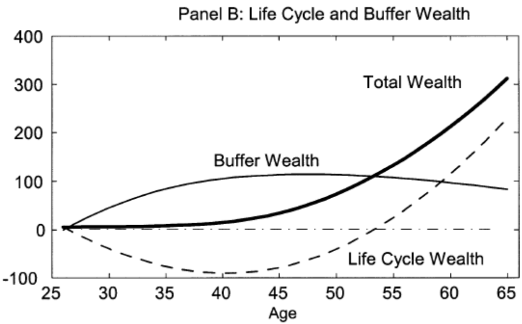


FIGURE 7.—The role of risk in saving and wealth accumulation.

Outline

1 Introduction

2 Stochastic DP

3 Structural Estimation

Buffer-stock model (Deaton-Carroll) Bellman equation

- **Simplest version** of the buffer-stock model is

$$
\begin{aligned}
V_t(M_t, P_t) &= \max_{C_t} \frac{C_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[V_{t+1}(M_{t+1}, P_{t+1}) \right] \\
&\text{s.t.} \\
A_t &= M_t - C_t && \text{(assets)} \\
M_{t+1} &= R A_t + Y_{t+1} && \text{(resources/cash-on-hand)} \\
Y_{t+1} &= P_{t+1} \xi_{t+1} && \text{(income)} \\
P_{t+1} &= G P_t \psi_{t+1} && \text{(perm. income)} \\
A_t &\geq 0, \forall t && \text{(no borrowing)}
\end{aligned}
$$

where $\mathbb{E}_t \left[V_{t+1}(M_{t+1}, P_{t+1}) \right] = \mathbb{E} \left[V_{t+1}(M_{t+1}, P_{t+1}) | M_t, P_t, C_t \right]$ are expectations over perm. and trans. income shocks,

$$\log \xi_{t+1} \sim \mathcal{N}(\mu_\xi, \sigma_\xi^2), \; \log \psi_{t+1} \sim \mathcal{N}(\mu_\psi, \sigma_\psi^2)$$

Buffer-stock model (Deaton-Carroll) Bellman equation

- **Simplest version** of the buffer-stock model is

$$
\begin{aligned}
V_t(M_t, P_t) &= \max_{C_t} \frac{C_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[V_{t+1}(M_{t+1}, P_{t+1}) \right] \\
&\text{s.t.} \\
A_t &= M_t - C_t && \text{(assets)} \\
M_{t+1} &= RA_t + Y_{t+1} && \text{(resources/cash-on-hand)} \\
Y_{t+1} &= P_{t+1}\xi_{t+1} && \text{(income)} \\
P_{t+1} &= GP_t\psi_{t+1} && \text{(perm. income)} \\
A_t &\geq 0, \forall t && \text{(no borrowing)}
\end{aligned}
$$

where $\mathbb{E}_t \left[V_{t+1}(M_{t+1}, P_{t+1}) \right] = \mathbb{E} \left[V_{t+1}(M_{t+1}, P_{t+1}) | M_t, P_t, C_t \right]$ are expectations over perm. and trans. income shocks,

$$\log \xi_{t+1} \sim \mathcal{N}(\mu_\xi, \sigma_\xi^2), \ \log \psi_{t+1} \sim \mathcal{N}(\mu_\psi, \sigma_\psi^2)$$

- Gourinchas and Parker (2002): "natural" borrowing constraint. mass-point at zero in trans. income shock distribution, ξ_{t+1}

Buffer-stock model (Deaton-Carroll) Bellman equation

- **Last period:** Everything is consumed,

$$C_T^\star(M_T, P_T) = M_T$$
$$V_T(M_T, P_T) = \frac{M_T^{1-\rho}}{1-\rho}$$

Buffer-stock model (Deaton-Carroll) Bellman equation

- **Last period:** Everything is consumed,

$$C_T^{\star}(M_T, P_T) = M_T$$
$$V_T(M_T, P_T) = \frac{M_T^{1-\rho}}{1-\rho}$$

- **Gourinchas and Parker (2002)**: Retirement-periods
  Assumes a linear post-retirement value (w. $P_{T+1} = P_T$)

$$V_{T+1}(M_{T+1}, P_{T+1}) = \kappa \cdot (M_{T+1} + h \cdot P_{T+1})$$

  Motivated by a deterministic perfect credit market solution (estimate κ and h, through γ_0 and γ_1 – see e.g. Jørgensen and Tô, 2020)
  - They also allow for time-varying taste-shifters, $v_t(Z_t)$.

Normalization I

- **Defining** $c_t \equiv C_t / P_t$, $m_t \equiv M_t / P_t$ etc. implies

$$
\begin{aligned}
A_t &= M_t - C_t \\
A_t / P_t &= M_t / P_t - C_t / P_t \\
a_t &= m_t - c_t
\end{aligned}
$$

Normalization I

- **Defining** $c_t \equiv C_t / P_t$, $m_t \equiv M_t / P_t$ etc. implies

$$
\begin{aligned}
A_t &= M_t - C_t \\
A_t / P_t &= M_t / P_t - C_t / P_t \\
a_t &= m_t - c_t
\end{aligned}
$$

and state transition

$$
\begin{aligned}
M_{t+1} &= RA_t + Y_{t+1} \\
M_{t+1} / P_{t+1} &= RA_t / P_{t+1} + Y_{t+1} / P_{t+1} \\
m_{t+1} &= R a_t P_t / P_{t+1} + \xi_{t+1} \\
m_{t+1} &= \frac{R}{G\psi_{t+1}} a_t + \xi_{t+1}
\end{aligned}
$$

The **adjustment factor** $\frac{1}{G\psi_{t+1}}$ is due to changes in permanent income

Normalization II

- **Defining** $v_t(m_t) = V_t(M_t, P_t)/P_t^{1-\rho}$ implies

$$
\begin{aligned}
V_t(M_t, P_t) &= \max_{C_t} \frac{C_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[V_{t+1}(M_{t+1}, P_{t+1}) \right] \\
&= \max_{c_t} \frac{(c_t P_t)^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[V_{t+1}(M_{t+1}, P_{t+1}) \right] \Leftrightarrow \\
V_t(M_t, P_t)/P_t^{1-\rho} &= \max_{c_t} \frac{(c_t P_t)^{1-\rho}/P_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[V_{t+1}(M_{t+1}, P_{t+1})/P_t^{1-\rho} \right] \Leftrightarrow \\
v_t(m_t) &= \max_{c_t} \frac{{c_t}^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \Big[\underbrace{V_{t+1}(M_{t+1}, P_{t+1})/P_{t+1}^{1-\rho}}_{=v_{t+1}(m_{t+1})} \cdot \underbrace{P_{t+1}^{1-\rho}/P_t^{1-\rho}}_{=(G\psi_{t+1})^{1-\rho}} \Big] \\
&= \max_{c_t} \frac{{c_t}^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[(G\psi_{t+1})^{1-\rho} v_{t+1}(m_{t+1}) \right]
\end{aligned}
$$

Bellman equation in ratio form

$$
\begin{aligned}
v_t(m_t) &= \max_{c_t} \frac{c_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[(G\psi_{t+1})^{1-\rho} v_{t+1}(m_{t+1}) \right] \\
&\text{s.t.} \\
a_t &= m_t - c_t \\
m_{t+1} &= \frac{1}{G\psi_{t+1}} R a_t + \xi_{t+1} \\
a_t &\geq 0
\end{aligned}
$$

- **Benefit:** Dimensionality of state space reduced, $2 \rightarrow 1$.
  Can this always be done?

Bellman equation in ratio form

$$
\begin{aligned}
v_t(m_t) &= \max_{c_t} \frac{c_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[(G\psi_{t+1})^{1-\rho} v_{t+1}(m_{t+1}) \right] \\
&\text{s.t.} \\
a_t &= m_t - c_t \\
m_{t+1} &= \frac{1}{G\psi_{t+1}} R a_t + \xi_{t+1} \\
a_t &\geq 0
\end{aligned}
$$

- **Benefit:** Dimensionality of state space reduced, $2 \to 1$.
  Can this always be done?
- No... Uses that utility is homothetic (budget constraint also important)

$$
V_T(M_T, P_T) = \frac{M_T^{1-\rho}}{1-\rho} = \frac{(m_T P_T)^{1-\rho}}{1-\rho} = \frac{m_T{}^{1-\rho}}{1-\rho} P_T^{1-\rho}
$$

such that $v_T(m_T) = V_T(M_T, P_T)/P_T^{1-\rho}$ holds!

Solving the model: Numerical Integration

- Solved by **backwards induction**
  Terminal period:

$$v_T(m_T) = \frac{m_T^{1-\rho}}{1-\rho}$$

For $t < T$:

$$v_t(m_t) \quad = \quad \max_{c_t} \frac{c_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[(G\psi_{t+1})^{1-\rho} v_{t+1}(m_{t+1}) \right]$$

Solving the model: Numerical Integration

- Solved by **backwards induction**
  Terminal period:
  $$v_T(m_T) = \frac{m_T^{1-\rho}}{1-\rho}$$
  For $t < T$:
  $$v_t(m_t) \;=\; \max_{c_t} \frac{c_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[(G\psi_{t+1})^{1-\rho} v_{t+1}(m_{t+1}) \right]$$
- **How to evaluate expectations?**
  $$\mathbb{E}_t[\bullet] = \int_{\psi_{t+1}} \int_{\xi_{t+1}} \left[(G\psi_{t+1})^{1-\rho} v_{t+1}(m_{t+1}) \right] f(d\psi_{t+1}, d\xi_{t+1})$$

Solving the model: Numerical Integration

- Solved by **backwards induction**
  Terminal period:

$$v_T(m_T) = \frac{m_T^{1-\rho}}{1-\rho}$$

  For $t < T$:

$$v_t(m_t) \quad = \quad \max_{c_t} \frac{c_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[(G\psi_{t+1})^{1-\rho} v_{t+1}(m_{t+1}) \right]$$

- **How to evaluate expectations?**

$$\mathbb{E}_t[\bullet] = \int_{\psi_{t+1}} \int_{\xi_{t+1}} \left[(G\psi_{t+1})^{1-\rho} v_{t+1}(m_{t+1}) \right] f(d\psi_{t+1}, d\xi_{t+1})$$

- **Numerical Integration**: Discretize into sum (Gauss-Hermite)

$$\mathbb{E}_t[\bullet] \approx \sum_{j=1}^{J} \sum_{k=1}^{K} [(G\psi^{(j)})^{1-\rho} v_{t+1}(m^{(j,k)})] \omega_j \omega_k$$

  and interpolate $v_{t+1}(\bullet)$ for values $m^{(j,k)} = \frac{1}{G\psi^{(j)}} R a_t + \xi^{(k)}$ of $\overrightarrow{m}$ grid.

Outline

Structural vs. Reduced-Form Estimation

- **Critique** of structural estimation: Requires many assumptions

Structural vs. Reduced-Form Estimation

- **Critique** of structural estimation: Requires many assumptions
- **Alternative:** Estimate reduced-form equations “derived” from model

Structural vs. Reduced-Form Estimation

- **Critique** of structural estimation: Requires many assumptions
- **Alternative:** Estimate reduced-form equations “derived” from model
- **My (and others) claim:** To turn reduced form parameter estimates into policy advice *requires a lot of assumptions*

  > *“All econometric work relies heavily on a priori assumptions. The main difference between structural and experimental (or “atheoretic”) approaches is not in the number of assumptions but the extent to which they are made explicit.” (Keane, 2010)*

Structural vs. Reduced-Form Estimation

- **Critique** of structural estimation: Requires many assumptions
- **Alternative:** Estimate reduced-form equations "derived" from model
- **My (and others) claim:** To turn reduced form parameter estimates into policy advice *requires a lot of assumptions*

  > *"All econometric work relies heavily on a priori assumptions. The main difference between structural and experimental (or "atheoretic") approaches is not in the number of assumptions but the extent to which they are made explicit." (Keane, 2010)*

- **Benefit of models:**
  1. Ensure *consistent* world view
  2. Assumptions are clear: Better models are well defined.
  3. Hopefully "deep" policy-invariant parameters (Lucas critique).

Structural vs. Reduced-Form Estimation

- **Critique** of structural estimation: Requires many assumptions
- **Alternative:** Estimate reduced-form equations "derived" from model
- **My (and others) claim:** To turn reduced form parameter estimates into policy advice *requires a lot of assumptions*

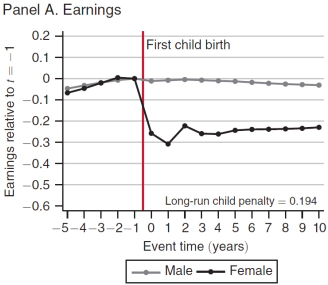
  > *"All econometric work relies heavily on a priori assumptions. The main difference between structural and experimental (or "atheoretic") approaches is not in the number of assumptions but the extent to which they are made explicit." (Keane, 2010)*

- **Benefit of models:**
  1. Ensure *consistent* world view
  2. Assumptions are clear: Better models are well defined.
  3. Hopefully "deep" policy-invariant parameters (Lucas critique).
- **Frontier:** Use exogenous variation to estimate structural model.

Structural vs. Reduced-Form Estimation

- **Example:** Event-studies (child-birth, Kleven, Landais and Søgaard, 2019)

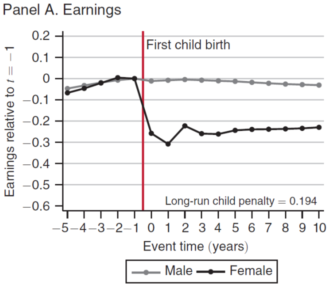
- **Reduced-form** to be *causal*: "statistical" assumptions
    - No self-selection (timing)
    - No anticipation effects.
    - Parallel trends.



Structural vs. Reduced-Form Estimation

- **Example:** Event-studies (child-birth, Kleven, Landais and Søgaard, 2019)

  - **Reduced-form** to be *causal*: "statistical" assumptions
    - No self-selection (timing)
    - No anticipation effects.
    - Parallel trends.



- **A model can allow** for these assumptions to be violated
  But only through the chosen functional forms and mechanisms
  - "Economic" assumptions
  - Easier to debate and approve upon (?)

Structural Estimation

- We know how to **solve dynamic programming models**
- **How can we estimate them?** We need
  1. Data on (some) *states*
  2. Data on (some) *choices*

Structural Estimation

- We know how to **solve dynamic programming models**
- **How can we estimate them?** We need
    1. Data on (some) *states*
    2. Data on (some) *choices*
- Two **standard approaches**
    1. Maximum likelihood (ML)
    2. General Method of Moments (GMM)

Structural Estimation

- We know how to **solve dynamic programming models**
- **How can we estimate them?** We need
  1. Data on (some) *states*
  2. Data on (some) *choices*
- Two **standard approaches**
  1. Maximum likelihood (ML)
  2. General Method of Moments (GMM)
- **Simulated versions:** (integrate over unobserved states)
  1. Maximum Simulated Likelihood (MSL, SML)
  2. Method of Simulated Moments (MSM, SMM, **SMD**)

Structural Estimation

- We know how to **solve dynamic programming models**
- **How can we estimate them?** We need
  1. Data on (some) *states*
  2. Data on (some) *choices*
- Two **standard approaches**
  1. Maximum likelihood (ML)
  2. General Method of Moments (GMM)
- **Simulated versions:** (integrate over unobserved states)
  1. Maximum Simulated Likelihood (MSL, SML)
  2. Method of Simulated Moments (MSM, SMM, **SMD**)
- **Example model:** Life-cycle buffer-stock model
  - States: M_{it}, P_{it}
  - Choice: C_{it}
- **Parameters** to estimate: $\theta = \{\beta, \rho\}$
  - Calibration: G, σ_ψ, σ_ξ, R, and λ ("known")

Simulated Method of Moments (SMM/SMD)

- $\Lambda^d = \frac{1}{N}\sum_{i=1}^{N} \Lambda_i^d$ are some **moments in the data**
  Could be avg., var, cov, regression-coefs, etc.

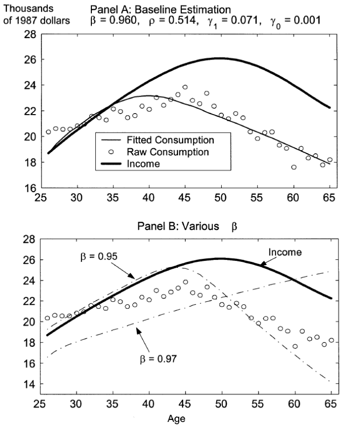


FIGURE 5.—The fitted consumption profile.

Simulated Method of Moments (SMM/SMD)

- $\Lambda^d = \frac{1}{N}\sum_{i=1}^{N} \Lambda_i^d$ are some **moments in the data**
  Could be avg., var, cov, regression-coefs, etc.
- $\Lambda^m(\theta)$ are the **same moments** calculated on **simulated data**

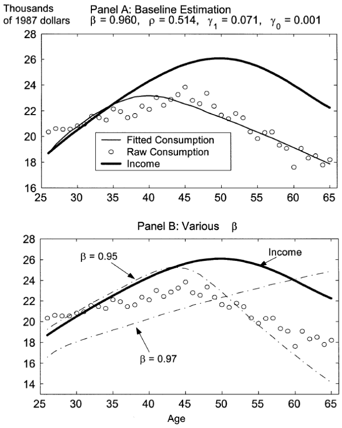


FIGURE 5.—The fitted consumption profile.

Simulated Method of Moments (SMM/SMD)

- $\Lambda^d = \frac{1}{N}\sum_{i=1}^{N} \Lambda_i^d$ are some **moments in the data** Could be avg., var, cov, regression-coefs, etc.
- $\Lambda^m(\theta)$ are the **same moments** calculated on **simulated data**
- The **difference** is then

$$g(\theta) = \Lambda^d - \Lambda^m(\theta)$$

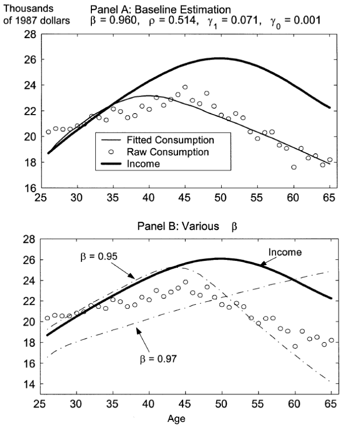


FIGURE 5.—The fitted consumption profile.

Simulated Method of Moments (SMM/SMD)

- $\Lambda^d = \frac{1}{N}\sum_{i=1}^{N} \Lambda_i^d$ are some **moments in the data**
  Could be avg., var, cov, regression-coefs, etc.
- $\Lambda^m(\theta)$ are the **same moments** calculated on **simulated data**
- The **difference** is then

$$g(\theta) = \Lambda^d - \Lambda^m(\theta)$$

- **SMM** then is

$$\hat{\theta} = \arg\min_{\theta} g(\theta)' W g(\theta)$$

where W is **weighting matrix**.

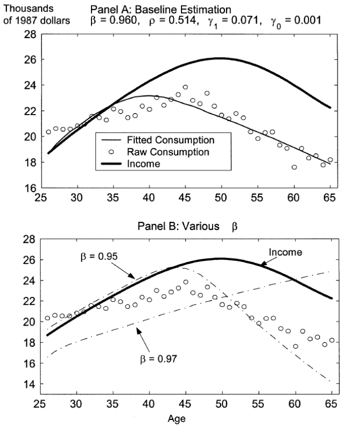


FIGURE 5.—The fitted consumption profile.

Weighting Matrix, W

- Common weighting matrices, W, are
  (should be positive-definite)

  1. **Theoretically optimal**
     Inverse of covariance matrix of empirical moments
     Can cause problems in finite samples
  2. **Identity, I**
     Equal weighting.
     Does not take level-differences out of moments
  3. **Diagonal matrix** with *inverse* of empirical moment *variances*
     Removes "level" differences.
     Scales with uncertainty about empirical moments
     Popular
  4. **Freely chosen**
     Focus on fitting some specific dimensions of the data

Estimation experiment

1. **Solve** the buffer-stock model and **simulate** a full panel
2. Construct a **data set** from the simulated data
3. Try to **estimate** $\theta = \{\beta, \rho\}$
   using as moments the **average wealth for each age between 40 and 55**
   $\Lambda^d = (A_{40}, A_{41}, \ldots, A_{55})$

Estimation experiment

1. **Solve** the buffer-stock model and **simulate** a full panel
2. Construct a **data set** from the simulated data
3. Try to **estimate** $\theta = \{\beta, \rho\}$
   using as moments the **average wealth for each age between 40 and 55**
   $\Lambda^d = (A_{40}, A_{41}, \ldots, A_{55})$

- I will now describe how to calculate the objective function

$$Q(\theta) = \left(\Lambda^d - \Lambda^m(\theta)\right)' W \left(\Lambda^d - \Lambda^m(\theta)\right)$$

  for a given value of θ.
- This function should then be minimized to get

$$\hat{\theta} = \arg\min_{\theta} Q(\theta)$$

Implementation, $\hat{\theta}_{MSM} = \arg\min_{\theta} Q(\theta)$

1. Solve model to get $c_t^{\star}(m;\theta)$ for all t on a grid of m (2-dim array)

Implementation, $\hat{\theta}_{MSM} = \arg\min_{\theta} Q(\theta)$

1. Solve model to get $c^{\star}_t(m;\theta)$ for all t on a grid of m (2-dim array)
2. For $s = 1, \dots, S$:
   2.1 Simulate N agents for T periods to get

$$
\begin{aligned}
C^{(s)}_{it}(\theta) &= P^{(s)}_{it} \cdot \check{c}^{\star}_t(M^{(s)}_{it}(\theta)/P^{(s)}_{it};\theta) \\
M^{(s)}_{it}(\theta) &= RA^{(s)}_{it-1}(\theta) + Y^{(s)}_{it} \\
A^{(s)}_{it-1}(\theta) &= M^{(s)}_{it-1}(\theta) - C^{(s)}_{it-1}(\theta) \\
Y^{(s)}_{it} &= P^{(s)}_{it}\xi^{(s)}_{it} \\
P^{(s)}_{it} &= GP^{(s)}_{it-1}\psi^{(s)}_{it}
\end{aligned}
$$

   for some initial A_{i0} and P_{i0} and draws of **?**

Implementation, $\hat{\theta}_{MSM} = \arg\min_{\theta} Q(\theta)$

1. Solve model to get $c_t^{\star}(m;\theta)$ for all t on a grid of m (2-dim array)
2. For $s = 1, \ldots, S$:
   2.1 Simulate N agents for T periods to get

$$\begin{aligned}
C_{it}^{(s)}(\theta) &= P_{it}^{(s)} \cdot \check{c}_t^{\star}(M_{it}^{(s)}(\theta)/P_{it}^{(s)};\theta) \\
M_{it}^{(s)}(\theta) &= RA_{it-1}^{(s)}(\theta) + Y_{it}^{(s)} \\
A_{it-1}^{(s)}(\theta) &= M_{it-1}^{(s)}(\theta) - C_{it-1}^{(s)}(\theta) \\
Y_{it}^{(s)} &= P_{it}^{(s)}\xi_{it}^{(s)} \\
P_{it}^{(s)} &= GP_{it-1}^{(s)}\psi_{it}^{(s)}
\end{aligned}$$

for some initial A_{i0} and P_{i0} and draws of $\xi_{it}^{(s)}$ and $\psi_{it}^{(s)}$.

Implementation, $\hat{\theta}_{MSM} = \arg\min_{\theta} Q(\theta)$

1. Solve model to get $c_t^{\star}(m;\theta)$ for all t on a grid of m (2-dim array)
2. For $s = 1, \dots, S$:
   2.1 Simulate N agents for T periods to get

$$
\begin{aligned}
C_{it}^{(s)}(\theta) &= P_{it}^{(s)} \cdot \check{c}_t^{\star}(M_{it}^{(s)}(\theta)/P_{it}^{(s)};\theta) \\
M_{it}^{(s)}(\theta) &= RA_{it-1}^{(s)}(\theta) + Y_{it}^{(s)} \\
A_{it-1}^{(s)}(\theta) &= M_{it-1}^{(s)}(\theta) - C_{it-1}^{(s)}(\theta) \\
Y_{it}^{(s)} &= P_{it}^{(s)}\xi_{it}^{(s)} \\
P_{it}^{(s)} &= GP_{it-1}^{(s)}\psi_{it}^{(s)}
\end{aligned}
$$

   for some initial A_{i0} and P_{i0} and draws of $\xi_{it}^{(s)}$ and $\psi_{it}^{(s)}$.

   2.2 Calculate moments using simulated data, $\Lambda_s(\theta) = \{\frac{1}{N}\sum_{i=1}^{N} A_{it}^{(s)}(\theta)\}_{t=40}^{55}$

Implementation, $\hat{\theta}_{MSM} = \arg\min_{\theta} Q(\theta)$

1. Solve model to get $c_t^{\star}(m;\theta)$ for all t on a grid of m (2-dim array)
2. For $s = 1, \ldots, S$:
   2.1 Simulate N agents for T periods to get

$$
\begin{aligned}
C_{it}^{(s)}(\theta) &= P_{it}^{(s)} \cdot \check{c}_t^{\star}(M_{it}^{(s)}(\theta)/P_{it}^{(s)};\theta) \\
M_{it}^{(s)}(\theta) &= RA_{it-1}^{(s)}(\theta) + Y_{it}^{(s)} \\
A_{it-1}^{(s)}(\theta) &= M_{it-1}^{(s)}(\theta) - C_{it-1}^{(s)}(\theta) \\
Y_{it}^{(s)} &= P_{it}^{(s)}\xi_{it}^{(s)} \\
P_{it}^{(s)} &= GP_{it-1}^{(s)}\psi_{it}^{(s)}
\end{aligned}
$$

   for some initial A_{i0} and P_{i0} and draws of $\xi_{it}^{(s)}$ and $\psi_{it}^{(s)}$.

   2.2 Calculate moments using simulated data, $\Lambda_s(\theta) = \{\frac{1}{N}\sum_{i=1}^{N} A_{it}^{(s)}(\theta)\}_{t=40}^{55}$
3. Calculate the objective function with $\Lambda^m(\theta) = \frac{1}{S}\sum_{s=1}^{S}\Lambda_s(\theta)$

$$
Q(\theta) = \left(\Lambda^d - \Lambda^m(\theta)\right)' W \left(\Lambda^d - \Lambda^m(\theta)\right)
$$

Alt. Implementation, $\hat{\theta}_{MSM} = \arg\min_\theta Q(\theta)$

1. Solve model to get $c_t^\star(m;\theta)$ for all t on a grid of m (2-dim array)

Alt. Implementation, $\hat{\theta}_{MSM} = \arg\min_\theta Q(\theta)$

1. Solve model to get $c_t^\star(m;\theta)$ for all t on a grid of m (2-dim array)
2. Simulate $\tilde{S} = SN$ agents for T periods to get

$$
\begin{aligned}
C_t^{(s)}(\theta) &= P_t^{(s)} \cdot \check{c}_t^\star(M_i^{(s)}(\theta)/P_t^{(s)};\theta) \\
M_t^{(s)}(\theta) &= RA_{t-1}^{(s)}(\theta) + Y_t^{(s)} \\
A_{t-1}^{(s)}(\theta) &= M_{t-1}^{(s)}(\theta) - C_{t-1}^{(s)}(\theta) \\
Y_t^{(s)} &= P_t^{(s)}\xi_t^{(s)} \\
P_t^{(s)} &= GP_{t-1}^{(s)}\psi_t^{(s)}
\end{aligned}
$$

for some initial A_0 and P_0 and draws of $\xi_t^{(s)}$ and $\psi_t^{(s)}$.

Alt. Implementation, $\hat{\theta}_{MSM} = \arg\min_\theta Q(\theta)$

1. Solve model to get $c_t^\star(m;\theta)$ for all t on a grid of m (2-dim array)
2. Simulate $\tilde{S} = SN$ agents for T periods to get

$$
\begin{aligned}
C_t^{(s)}(\theta) &= P_t^{(s)} \cdot \check{c}_t^\star(M_i^{(s)}(\theta)/P_t^{(s)};\theta) \\
M_t^{(s)}(\theta) &= RA_{t-1}^{(s)}(\theta) + Y_t^{(s)} \\
A_{t-1}^{(s)}(\theta) &= M_{t-1}^{(s)}(\theta) - C_{t-1}^{(s)}(\theta) \\
Y_t^{(s)} &= P_t^{(s)}\xi_t^{(s)} \\
P_t^{(s)} &= GP_{t-1}^{(s)}\psi_t^{(s)}
\end{aligned}
$$

   for some initial A_0 and P_0 and draws of $\xi_t^{(s)}$ and $\psi_t^{(s)}$.
3. Calculate simulated moments, $\Lambda^m(\theta) = \{\frac{1}{\tilde{S}}\sum_{s=1}^{\tilde{S}} A_t^{(s)}(\theta)\}_{t=40}^{55}$ now

Alt. Implementation, $\hat{\theta}_{MSM} = \arg\min_\theta Q(\theta)$

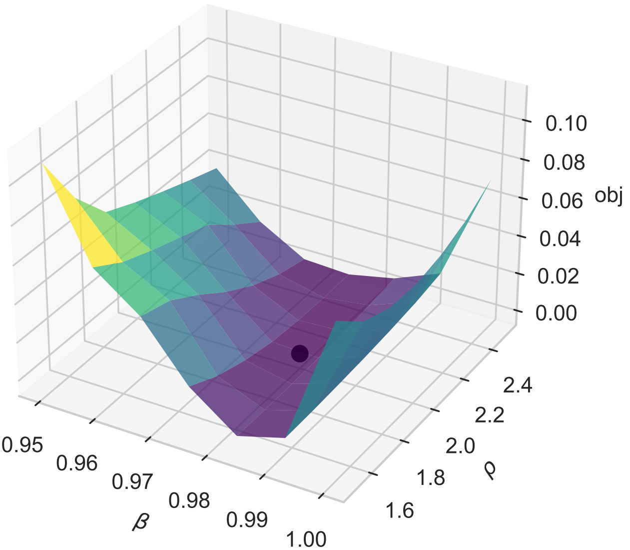
1. Solve model to get $c_t^\star(m;\theta)$ for all t on a grid of m (2-dim array)
2. Simulate $\tilde{S} = SN$ agents for T periods to get

$$
\begin{aligned}
C_t^{(s)}(\theta) &= P_t^{(s)} \cdot \check{c}_t^\star(M_i^{(s)}(\theta)/P_t^{(s)};\theta) \\
M_t^{(s)}(\theta) &= RA_{t-1}^{(s)}(\theta) + Y_t^{(s)} \\
A_{t-1}^{(s)}(\theta) &= M_{t-1}^{(s)}(\theta) - C_{t-1}^{(s)}(\theta) \\
Y_t^{(s)} &= P_t^{(s)}\xi_t^{(s)} \\
P_t^{(s)} &= GP_{t-1}^{(s)}\psi_t^{(s)}
\end{aligned}
$$

   for some initial A_0 and P_0 and draws of $\xi_t^{(s)}$ and $\psi_t^{(s)}$.
3. Calculate simulated moments, $\Lambda^m(\theta) = \{\frac{1}{\tilde{S}}\sum_{s=1}^{\tilde{S}} A_t^{(s)}(\theta)\}_{t=40}^{55}$ now
4. Calculate the objective function

$$
Q(\theta) = \Big(\Lambda^d - \Lambda^m(\theta)\Big)' W \Big(\Lambda^d - \Lambda^m(\theta)\Big)
$$

Buffer-stock: MSM



Indirect inference / minimum distance

- Many different names for very similar approaches
  - McFadden (1989): Method of Simulated Moments (MSM)
  - Duffie and Singleton (1993): Simulated Minimum Distance (SMD)
  - Gouriéroux, Monfort and Renault (1993) + Smith (1993): Indirect Inference (II)

Indirect inference / minimum distance

- Many different names for very similar approaches
  - McFadden (1989): Method of Simulated Moments (MSM)
  - Duffie and Singleton (1993): Simulated Minimum Distance (SMD)
  - Gouriéroux, Monfort and Renault (1993) + Smith (1993): Indirect Inference (II)
- SMD/II rely on an **auxillary statistical model**
  - Let Λ^d be the parameters of the auxillary model when estimated on the *actual* data
  - Let $\Lambda_s(\theta)$ be the parameters of the auxillary model when estimated on *simulated* data
- **Note:** The auxillary statistical model is *misspecified* and its parameters are thus typically *not interpretable*

Simulation Pitfalls

- **FIX the seed (or draws!)**
- **Flat** objective function!
  - Discrete choices: Taking a mean of an **indicator function**
- **Gradient** based numerical optimization will likely FAIL!
  - Use, e.g., `scipy.optimize.minimize(fun , method='Nelder-Mead')` (Nelder-Mead)
  - Or some smoothing device (e.g. Logit)
- As $N, S \rightarrow \infty$ this problem vanishes
- The problem is also less severe around θ_0
- Continuous outcomes do not have this problem

Asymptotics

- **MSM** is **consistent** and **asymptotically normal** under standard assumptions
$$\sqrt{N}(\hat{\theta} - \theta_0) \rightarrow \mathcal{N}(0, (1 + S^{-1})V)$$
where θ_0 is vector of true parameters
- **Standard formulas for V:**
$$V = (G'WG)^{-1} G'W\Omega W'G(G'WG)^{-1}$$
where $G = -\frac{\partial \Lambda^m(\theta)}{\partial \theta}$ is the Jacobian of the objective function.
$\Omega = Var(\Lambda_i^d)$ is the variance of the (individual) moments in the data.
Remember: *Standard errors are large if large changes in θ imply small changes in the objective function*

Identification

- **Is there enough variation in the data to identify θ?**
  Very hard to *prove* anything because the model is typically strongly non-linear

-

Identification

- **Is there enough variation in the data to identify θ?**
  Very hard to *prove* anything because the model is typically strongly non-linear
- **MSM:** At least the same number of moments as parameters

Identification

- **Is there enough variation in the data to identify θ?**
  Very hard to *prove* anything because the model is typically strongly non-linear
- **MSM:** At least the same number of moments as parameters
- "Informativeness of estimation moments" (Honoré, Jørgensen and de Paula, 2020)
  "How much does the variance of θ increase if we drop a (set of) moments?"
-

Identification

- **Is there enough variation in the data to identify θ?**
  Very hard to *prove* anything because the model is typically strongly non-linear
- **MSM:** At least the same number of moments as parameters
- "Informativeness of estimation moments" (Honoré, Jørgensen and de Paula, 2020)
  "How much does the variance of θ increase if we drop a (set of) moments?"
- **Graphical inspection is useful:** Plot the objective function in the neighborhood of the found optimum

-

Identification

- **Is there enough variation in the data to identify θ?**
  Very hard to *prove* anything because the model is typically strongly non-linear
- **MSM:** At least the same number of moments as parameters
- "Informativeness of estimation moments" (Honoré, Jørgensen and de Paula, 2020)
  "How much does the variance of θ increase if we drop a (set of) moments?"
- **Graphical inspection is useful:** Plot the objective function in the neighborhood of the found optimum
- **Problems:**
  1. The objective function might have multiple minima *(no global solver exists)*
  2. The objective function could be very flat in some directions *(increasing S might help)*
-

Robustness/Sensitivity

- **Curse of dimensionality and lack of identification**

  $\Rightarrow$ we cannot estimate all the parameters of the model

  $\Rightarrow$ *first step estimation/calibration is often necessary*

  1. Calculations on own data (e.g. exogenous processes)
  2. References to previous estimates
  3. Standard choices

Robustness/Sensitivity

- **Curse of dimensionality and lack of identification**
  $\Rightarrow$ we cannot estimate all the parameters of the model
  $\Rightarrow$ *first step estimation/calibration is often necessary*
  1. Calculations on own data (e.g. exogenous processes)
  2. References to previous estimates
  3. Standard choices
- **Robustness:** Can we vary the calibration choices without changing the result substantially?
  "Sensitivity to Calibration": (Jørgensen, forthcoming)
  "*How much does estimates of θ change when 1. step calibrations change?*"

Calibration vs. Estimation

- **Estimation** or **calibration:** What is the difference? (my take)

Calibration vs. Estimation

- **Estimation** or **calibration:** What is the difference?
(my take)

- **Estimation:** “systematic”
Use a solver to minimize a criteria function wrt. θ
Report standard errors on $\hat{\theta}$
Time-consuming!

Calibration vs. Estimation

- **Estimation** or **calibration:** What is the difference? (my take)
- **Estimation:** "systematic"
  Use a solver to minimize a criteria function wrt. θ
  Report standard errors on $\hat{\theta}$
  Time-consuming!
- **Calibration:** "hand-held"
  Use a (small) grid of values for θ
  to minimize some moment(s). Sometimes eyeballing, sequentially for each parameter.
  Do not report standard errors
  Less time-consuming!

Next Time

- **Next time:**
  Static and dynamic labor supply
  Recap for some + new stuff for most.

- **Literature:**
  Keane (2011, sections 1–5): "Labor Supply and Taxes: A Survey"
  - **Read** before lecture
  - **Reading guide:**
    Section 1: short Introduction
    Section 2: Optimal Taxation, Motivation. Skim fast.
    Section 3: Basic model. *Key, focus here*.
    Section 4: Econometric issues. Skim.
    Section 5: Roadmap of empirical literature. *Short, read*.
    (Remaining: empirical literature.)

References I

Carroll, C. D. (1992): "The Buffer-Stock Theory of Saving: Some Macroeconomic Evidence," *Brookings Papers on Economic Activity*, 2, 61–135.

Deaton, A. (1991): "Saving and Liquidity Constraints," *Econometrica*, 59(5), 1221–1248.

Duffie, D. and K. J. Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61(4), 929–952.

Gouriéroux, C., A. Monfort and E. Renault (1993): "Indirect Inference," *Journal of Applied Econometrics*, 8, 85–118.

Gourinchas, P.-O. and J. A. Parker (2002): "Consumption Over the Life Cycle," *Econometrica*, 70(1), 47–89.

References II

Honoré, B. E., T. H. Jørgensen and A. de Paula (2020): "The Informativeness of Estimation Moments," *Journal of Applied Econometrics*, 35(7), 797–813.

Jørgensen, T. and M. Tô (2020): "Robust Estimation of Finite Horizon Dynamic Economic Models," *Computational Economics*, 55(2), 499–509.

Jørgensen, T. H. (forthcoming): "Sensitivity to Calibrated Parameters," *Review of Economics and Statistics*.

Keane, M. P. (2010): "Structural vs. atheoretic approaches to econometrics," *Journal of Econometrics*, 156(1), 3–20.

——— (2011): "Labor Supply and Taxes: A Survey," *Journal of Economic Literature*, 49(4), 961–1075.

Kleven, H. J., C. Landais and J. E. Søgaard (2019): "Children and gender inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 11, 181–209.

References III

McFadden, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 47(5), 995–1026.

Smith, A. A. (1993): "Estimating nonlinear time-series models using simulated vector autoregressions," *Journal of Applied Econometrics*, 8, 63–84.