More on Structural Estimation

Dynamic Programming and Structural Econometrics #14

Bertel Schjerning

Reduced form or structural estimation

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

Reduced form or structural estimation

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

Reduced form or structural estimation

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

  > *"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, JoE 2012)*

Reduced form or structural estimation

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

  > *"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, JoE 2012)*

- **The beauty of models:**
  1. Ensure *consistent* world view
  2. Allow us to combine *heterogenous facts* and extrapolate from a myriad of past experiences
  3. Better models are clearly defined – even if we never find *the* true model we can make *progress*

Reduced form or structural estimation

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

  > *"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, JoE 2012)*

- **The beauty of models:**
  1. Ensure *consistent* world view
  2. Allow us to combine *heterogenous facts* and extrapolate from a myriad of past experiences
  3. Better models are clearly defined – even if we never find *the* true model we can make *progress*
- **Frontier:** Combine the two and use exogenous variation to estimate structural model.

The Lucas critique

- **The Lucas critique:** *Behavioral rules change with policy*
  $\Rightarrow$ policy advice can not rely on estimated behavioral rules
  $\Rightarrow$ we need to estimate *structural parameters*
  > *"Invariance of parameters in an economic model is not, of course, a property which can be assured in advance, but it seems reasonable to hope that neither tastes nor technology vary systematically with variations in counter-cyclical policies." (Lucas, 1977)*
- **Other stuff might be approximately invariant**
- **Rigourous microfoundations:**
  1. **Mathematically:** Based on (boundedly) rational behavior derived as a solution to a formal optimization problem
  2. **Economically:** The assumptions are realistic

Heavy duty econometrics

- Structural estimation requires a lot!
  1. solving a model *fast*
  2. having data to identify parameters
  3. setting up an estimation routine
  4. getting sensible estimates from that routine
  5. validate the model
  6. run meaningful counterfactuals
- Empirical work in term paper
  - Be modest with your ambitions and start simple

Example: Buffer-stock consumptions-savings model I

Bellman Equation

$$
\begin{aligned}
v_t(m_t) &= \max_{c_t} \left\{ \frac{c_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}_t \left[(GL_{t+1}\psi_{t+1})^{1-\rho} v_{t+1}(m_{t+1}) \right] \right\} \\
&\text{s.t.} \\
a_t &= m_t - c_t \\
m_{t+1} &= \frac{1}{GL_t \psi_{t+1}} R a_t + \xi_{t+1} \\
\xi_{t+1} &= \begin{cases} \mu & \text{with prob. } \pi \\ (\epsilon_{t+1} - \pi\mu)/(1-\pi) & \text{else} \end{cases} \\
a_t &\geq \lambda_t = \begin{cases} -\lambda & if t < T_R \\ 0 & if t \geq T_R \end{cases} \\
\psi_t &\sim \exp \mathcal{N}(-0.5\sigma_\psi^2, \sigma_\psi^2) \\
\epsilon_t &\sim \exp \mathcal{N}(-0.5\sigma_\xi^2, \sigma_\xi^2)
\end{aligned}
$$

Example: Buffer-stock consumptions-savings model II

$$
\begin{aligned}
Y_{t+1} &= \psi_{t+1} P_{t+1} \\
P_{t+1} &= G L_t P_t \psi_{t+1} \\
c_t &\equiv C_t / P_t \\
m_t &\equiv M_t / P_t \\
a_t &\equiv A_t / P_t \\
p_t &\equiv \ln(P_t) \\
y_t &\equiv \ln(Y_t)
\end{aligned}
$$

Structural estimation

- We know how to **solve dynamic programming models**
- We know how to **estimate dicrete choice models** (NFXP, CCP, NPL, MPEC, BBL)
- For estimation we need
    1. Data on (some) *states*
    2. Data on (some) *choices*

Structural estimation

- We know how to **solve dynamic programming models**
- We know how to **estimate dicrete choice models** (NFXP, CCP, NPL, MPEC, BBL)
- For estimation 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**
- We know how to **estimate dicrete choice models** (NFXP, CCP, NPL, MPEC, BBL)
- For estimation 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:**
  1. Maximum Simulated Likelihood (MSL, SML)
  2. Method of Simulated Moments (MSM, SMM)

Structural estimation

- We know how to **solve dynamic programming models**
- We know how to **estimate dicrete choice models** (NFXP, CCP, NPL, MPEC, BBL)
- For estimation 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:**
  1. Maximum Simulated Likelihood (MSL, SML)
  2. Method of Simulated Moments (MSM, SMM)
- **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")

Maximum likelihood estimation (MLE)

- Assume that observed log-consumption is contaminated with mean-zero i.i.d. normal **measurement error**

$$\epsilon_{it}(\theta) \equiv \log C_{it} - \log C_t^\star(M_{it}, P_{it}; \theta) \sim \mathcal{N}(0, \sigma_\xi^2)$$

Maximum likelihood estimation (MLE)

- Assume that observed log-consumption is contaminated with mean-zero i.i.d. normal **measurement error**

$$\epsilon_{it}(\theta) \equiv \log C_{it} - \log C^{\star}_t(M_{it}, P_{it}; \theta) \sim \mathcal{N}(0, \sigma^2_{\xi})$$

- The **“likelihood”** (pdf) of observing the data then is

$$\Pr(C, M, P|\theta) = \Pi_{i=1}^{N} \Pi_{t=1}^{T_d} \phi(\epsilon_{it}(\theta))$$

where $C = \{C_{it}\}_{1,1}^{N,T_d}$, $M = \{M_{it}\}_{1,1}^{N,T_d}$ and $P = \{P_{it}\}_{1,1}^{N,T_d}$ and

$$\phi(\epsilon_{it}) = \frac{1}{\sqrt{2\pi\sigma^2_{\xi}}} \exp\left(-\frac{{\epsilon_{it}}^2}{2\sigma^2_{\xi}}\right)$$

is the Gaussian density function

Maximum likelihood estimation (MLE)

- Assume that observed log-consumption is contaminated with mean-zero i.i.d. normal **measurement error**

$$\epsilon_{it}(\theta) \equiv \log C_{it} - \log C_t^\star(M_{it}, P_{it}; \theta) \sim \mathcal{N}(0, \sigma_\xi^2)$$

- The **"likelihood"** (pdf) of observing the data then is

$$\Pr(C, M, P|\theta) = \Pi_{i=1}^{N} \Pi_{t=1}^{T_d} \phi(\epsilon_{it}(\theta))$$

  where $C = \{C_{it}\}_{1,1}^{N,T_d}$, $M = \{M_{it}\}_{1,1}^{N,T_d}$ and $P = \{P_{it}\}_{1,1}^{N,T_d}$ and

$$\phi(\epsilon_{it}) = \frac{1}{\sqrt{2\pi\sigma_\xi^2}} \exp\left(-\frac{{\epsilon_{it}}^2}{2\sigma_\xi^2}\right)$$

  is the Gaussian density function

- **MLE** then is

$$\hat{\theta} = \arg\min_{\theta} -\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T_d} \log(\phi(\epsilon_{it}(\theta)))$$

Note: *We need to resolve the model for each new guess of* θ

Maximum Simulated Likelihood (MSL)

- MLE Requires that we observe *permanent* income, P_{it}!

Maximum Simulated Likelihood (MSL)

- MLE Requires that we observe *permanent* income, P_{it}!
- Integrate it out to get the marginal likelihood:

$$\begin{aligned}
\Pr(C, M|\theta) &= \mathbb{E}[\Pr(C, M, P|\theta)|M, \theta] \\
&= \int_{P_{T_d}} \int_{P_{T-1}} \cdots \int_{P_1} \Pr(C, M, P|\theta) dP_T dP_{T-1} \cdots dP_1
\end{aligned}$$

Maximum Simulated Likelihood (MSL)

- MLE Requires that we observe *permanent* income, P_{it}!
- Integrate it out to get the marginal likelihood:

$$\begin{aligned}\Pr(C, M|\theta) &= \mathbb{E}[\Pr(C, M, P|\theta)|M, \theta] \\ &= \int_{P_{T_d}} \int_{P_{T-1}} \cdots \int_{P_1} \Pr(C, M, P|\theta) dP_T dP_{T-1} \cdots dP_1\end{aligned}$$

- Drawing S draws of $P_1, P_2, \ldots, P_{T_d}$ could be used to get

$$\Pr\widehat{(C, M}|\theta) = \frac{1}{S}\sum_{s=1}^{S} \Pr(C, M, P^{(s)}|\theta)$$

Maximum Simulated Likelihood (MSL)

- MLE Requires that we observe *permanent* income, P_{it}!
- Integrate it out to get the marginal likelihood:

$$\begin{aligned}\Pr(C,M|\theta) &= \mathbb{E}[\Pr(C,M,P|\theta)|M,\theta] \\ &= \int_{P_{T_d}}\int_{P_{T-1}}\cdots\int_{P_1}\Pr(C,M,P|\theta)dP_T dP_{T-1}\cdots dP_1\end{aligned}$$

- Drawing S draws of $P_1, P_2, \ldots, P_{T_d}$ could be used to get

$$\Pr\widehat{(C,M}|\theta) = \frac{1}{S}\sum_{s=1}^{S}\Pr(C,M,P^{(s)}|\theta)$$

- and the MSL estimator is then

$$\hat{\theta} \quad = \quad \arg\min_{\theta} -\log\left(\Pr\widehat{(C,M}|\theta)\right)$$

Maximum Simulated Likelihood (MSL)

- MLE Requires that we observe *permanent* income, P_{it}!
- Integrate it out to get the marginal likelihood:

$$
\begin{aligned}
\Pr(C,M|\theta) &= \mathbb{E}[\Pr(C,M,P|\theta)|M,\theta] \\
&= \int_{P_{T_d}} \int_{P_{T-1}} \cdots \int_{P_1} \Pr(C,M,P|\theta) dP_T dP_{T-1} \cdots dP_1
\end{aligned}
$$

- Drawing S draws of $P_1, P_2, \ldots, P_{T_d}$ could be used to get

$$\Pr\widehat{(C,M}|\theta) = \frac{1}{S}\sum_{s=1}^{S} \Pr(C,M,P^{(s)}|\theta)$$

- and the MSL estimator is then

$$\hat{\theta} \quad = \quad \arg\min_{\theta} -\log\left(\Pr\widehat{(C,M}|\theta)\right)$$

- ***We would need many draws to approximate this T-dimensional integral***
  - *We may get a cosistent estimator of* $\Pr(C,M|\theta)$, *but not* $\log(\Pr(C,M|\theta)$ *(Jensen's Inequality)*
  - **MSL is inconsistent** for fixed S. The number of simulations S has to increase at least in the same rate the number of observations to achive consistency

Method of Simulated Moments (MSM)

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

Method of Simulated Moments (MSM)

- Let $\Lambda^d = \frac{1}{N}\sum_{i=1}^{N}\Lambda_i^d$ be some **moments in the data**
  - Could be avg., var, cov, regression-coefs, etc.
- Let $\Lambda_s(\theta)$ be the **same moments** calculated on data **simulated** from the model solved with parameters θ

Method of Simulated Moments (MSM)

- Let $\Lambda^d = \frac{1}{N}\sum_{i=1}^{N}\Lambda_i^d$ be some **moments in the data**
  - Could be avg., var, cov, regression-coefs, etc.
- Let $\Lambda_s(\theta)$ be the **same moments** calculated on data **simulated** from the model solved with parameters θ
- **MSM** then is

$$\hat{\theta} = \arg\min_{\theta}\left(\Lambda^d - \Lambda^m(\theta)\right)' W \left(\Lambda^d - \Lambda^m(\theta)\right)$$

where we average across simulations

$$\Lambda^m(\theta) \equiv \frac{1}{S}\sum_{s=1}^{S}\Lambda_s(\theta)$$

and W is a positive-definite **weighting matrix**.

Method of Simulated Moments (MSM)

- Let $\Lambda^d = \frac{1}{N}\sum_{i=1}^{N}\Lambda_i^d$ be some **moments in the data**
  - Could be avg., var, cov, regression-coefs, etc.
- Let $\Lambda_s(\theta)$ be the **same moments** calculated on data **simulated** from the model solved with parameters θ
- **MSM** then is

$$\hat{\theta} = \arg\min_{\theta} \left(\Lambda^d - \Lambda^m(\theta)\right)' W \left(\Lambda^d - \Lambda^m(\theta)\right)$$

  where we average across simulations

$$\Lambda^m(\theta) \equiv \frac{1}{S}\sum_{s=1}^{S}\Lambda_s(\theta)$$

  and W is a positive-definite **weighting matrix**.
- We still approximate the T_d-dimensional integral with S simulations but MSM is consistent for a fixed number of draws (Jensen's inequality does not kick in the same way)

Weighting matrix

- Typical choices are

    1. **Theoretically optimal** (Full vaiance-covariance matrix of moments)
       Can cause problems in finite samples
    2. **Diagonal matrix** with **inverse** of (e.g. bootstrapped) empirical **variances of the moments** (scaled appropriately)
    3. **Freely chosen** to focus on fitting some specific dimensions of the data

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)
  - Gourieroux, 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)
  - Gourieroux, 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 $N, S \to \infty$ this problem is ameliorated
- The problem is usually 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 are the 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*
- **Computational limitations:** To compute standard errors we need to compute derivative so model solution

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
- **Problems:**
  1. The objective function might have multiple minima
  2. The objective function could be very flat in some directions

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
- **Problems:**
  1. The objective function might have multiple minima
  2. The objective function could be very flat in some directions
- **Graphical inspection is useful:**
  - Investigate godness of fit - also for non-targeted moments
  - Plot the objective function in the neighborhood of the found optimum (vary a subset of 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
- **Problems:**
  1. The objective function might have multiple minima
  2. The objective function could be very flat in some directions
- **Graphical inspection is useful:**
  - Investigate godness of fit - also for non-targeted moments
  - Plot the objective function in the neighborhood of the found optimum (vary a subset of parameters)
- **Use more data**
  1. **Quantitatively:** More agents, more time periods
  2. **Qualitative:** New types of data, e.g natural experiments around policy changes

Estimation experiment

1. **Solve** the buffer-stock model and **simulate** a full panel
2. Construct a **data set** from the simulated data

   Likelihood: Log-consumption at age 45 with measurement error
   MSM: Average wealth for each age between 40 and 55
3. Try to **estimate** $\theta = \{\beta, \rho\}$

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

For Λ^d and a given value of θ, $Q(\theta)$:

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

For Λ^d and a given value of θ, $Q(\theta)$:

1. Solve model to get $c_t^\star(m;\theta)$ on a grid of m

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

For Λ^d and a given value of θ, $Q(\theta)$:

1. Solve model to get $c_t^{\star}(m;\theta)$ on a grid of m
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 **?**

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

For Λ^d and a given value of θ, $Q(\theta)$:

1. Solve model to get $c_t^\star(m;\theta)$ on a grid of m
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, $\hat{\theta}_{MSM} = \arg\min_{\theta} Q(\theta)$

For Λ^d and a given value of θ, $Q(\theta)$:

1. Solve model to get $c_t^{\star}(m;\theta)$ on a grid of m
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 the moments using this simulated data, $\Lambda_s(\theta)$ $(\{\frac{1}{N}\sum_{i=1}^{N} A_{it}^{(s)}(\theta)\}_{t=40}^{55})$

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

For Λ^d and a given value of θ, $Q(\theta)$:

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

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

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

   2.2 Calculate the moments using this 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, $\hat{\theta}_{MSM} = \arg\min_\theta Q(\theta)$

For Λ^d and a given value of θ, $Q(\theta)$:

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

For Λ^d and a given value of θ, $Q(\theta)$:

1. Solve model to get $c_t^\star(m;\theta)$ on a grid of m

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

For Λ^d and a given value of θ, $Q(\theta)$:

1. Solve model to get $c_t^\star(m;\theta)$ on a grid of m
2. Simulate S 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, $\hat{\theta}_{MSM} = \arg\min_\theta Q(\theta)$

For Λ^d and a given value of θ, $Q(\theta)$:

1. Solve model to get $c_t^\star(m;\theta)$ on a grid of m
2. Simulate S 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 the moments using this simulated data, $\Lambda^m(\theta)$ ($\{\frac{1}{S}\sum_{s=1}^S A_t^{(s)}(\theta)\}_{t=40}^{55}$ here)

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

For Λ^d and a given value of θ, $Q(\theta)$:

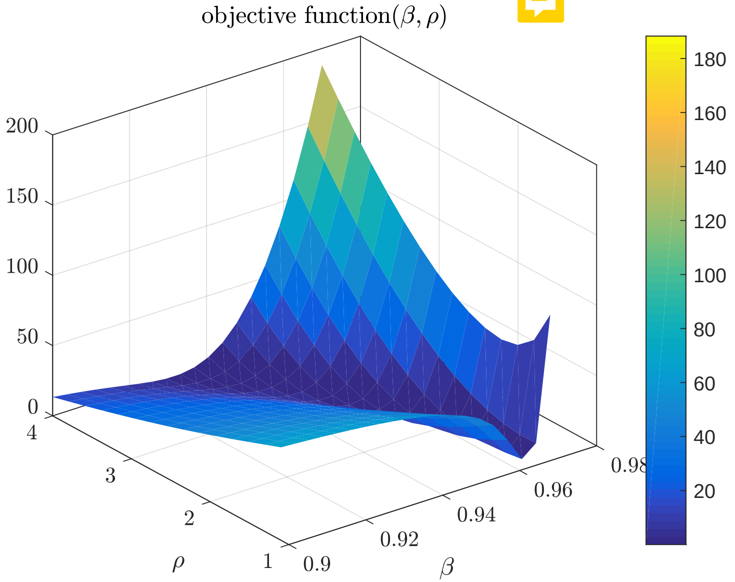
1. Solve model to get $c_t^\star(m;\theta)$ on a grid of m
2. Simulate S agents for T periods to get

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

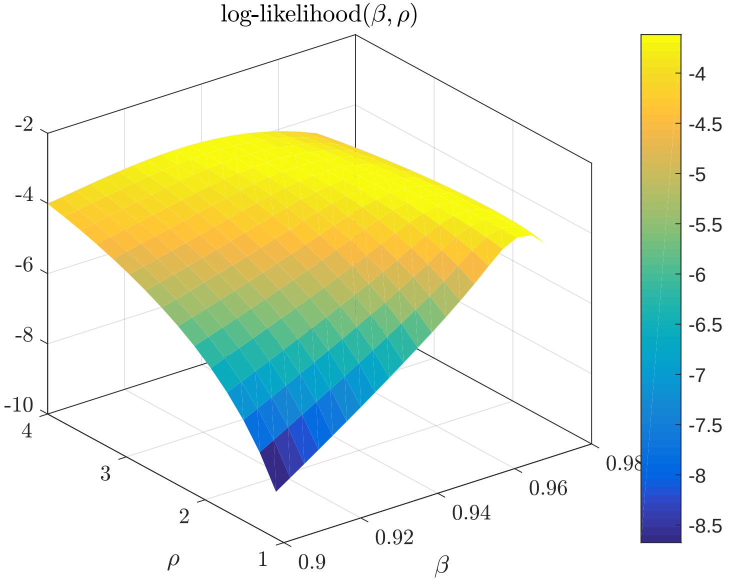
   for some initial A_0 and P_0 and draws of $\xi_t^{(s)}$ and $\psi_t^{(s)}$.
3. Calculate the moments using this simulated data, $\Lambda^m(\theta)$ ($\{\frac{1}{S}\sum_{s=1}^S A_t^{(s)}(\theta)\}_{t=40}^{55}$ here)
4. Calculate the objective function

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

Buffer-stock: MSM



Buffer-stock: Likelihood



Robustness

- **Curse of dimensionality and lack of identification**
  ⇒ somimes we cannot estimate all the parameters of the model
  ⇒ *first step calibration may be necessary*
  1. Calculations on own data (e.g. exogenous processes)
  2. References to previous estimates
  3. Standard choices

Robustness

- **Curse of dimensionality and lack of identification**
  $\Rightarrow$ somimes we cannot estimate all the parameters of the model
  $\Rightarrow$ *first step calibration may be 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?
  - **Or the opposite:** When does the result break down?

Robustness

- **Curse of dimensionality and lack of identification**
  $\Rightarrow$ somimes we cannot estimate all the parameters of the model
  $\Rightarrow$ *first step calibration may be 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?
  - **Or the opposite:** When does the result break down?
- **Calibration** is also important for
  1. Gaining intuition for how the model works
  2. Initial guesses for estimation algorithm

Examples

- **Gourinchas and Parker (2002):** First structural estimation of buffer-stock consumption model
  - **Method:** MSM with a lot of first stage calibrations
  - **Data:** Cross-sectional consumption data from CEX

Examples

- **Gourinchas and Parker (2002):** First structural estimation of buffer-stock consumption model
  - **Method:** MSM with a lot of first stage calibrations
  - **Data:** Cross-sectional consumption data from CEX
- **More examples - all estimated using MSM**
  1. **Druedahl and Jørgensen (Economic Journal, 2020): "Can Consumers Distinguish Persistent from Transitory Income Shocks?"**
     MSM Monte Carlo study: Pssoible to identify consumers' degree of information by using panel data on income and consumption
  2. Keane and Wasi (**Economic Journal, 20**16) : Labour Supply: The Roles of Human Capital and The Extensive Margin
  3. Iskhakov and Keane (Journal of Econometrics, 2020): "Effects of taxes and safety net pensions on life-cycle labor supply, savings and human capital: The case of Australia"
     3.1 We will look at this paper in detail in the next lecture (nice application of DC-EGM)

Gourinchas and Parker (2002) I

TABLE III

STRUCTURAL ESTIMATION RESULTS

MSM Estimation	Robust Weighting	Optimal Weighting
Discount Factor (β)	0.9598	0.9569
S.E.(A)	(0.0101)	
S.E.(B)	(0.0179)	(0.0150)
Discount Rate ($\beta^{-1}-1$)(%)	4.188	4.507
S.E.(A)	(1.098)	
S.E.(B)	(1.949)	(1.641)
Risk Aversion (ρ)	0.5140	1.3969
S.E.(A)	(0.1690)	
S.E.(B)	(0.1707)	(0.1137)
Retirement Rule:		
γ_0	0.0015	$5.68\,10^{-6}$
S.E.(A)	(3.84)	
S.E.(B)	(3.85)	(16.49)
γ_1	0.0710	0.0613
S.E.(A)	(0.1215)	
S.E.(B)	(0.1244)	(0.0511)
χ^2(A)	175.25	
χ^2(B)	174.10	185.67

Note: MSM estimation for entire group. Standard errors calculated without (A) and with (B) correction for first stage estimation. Cell size is 36,691 households. The last row reports a test of the overidentifying restrictions distributed as a Chi-squared with 36 degrees of freedom. The critical value at 5% is 50.71. Efficient estimates are calculated with a weighting matrix $\widehat{\Omega}$ computed from the robust estimates.

Gourinchas and Parker (2002) II

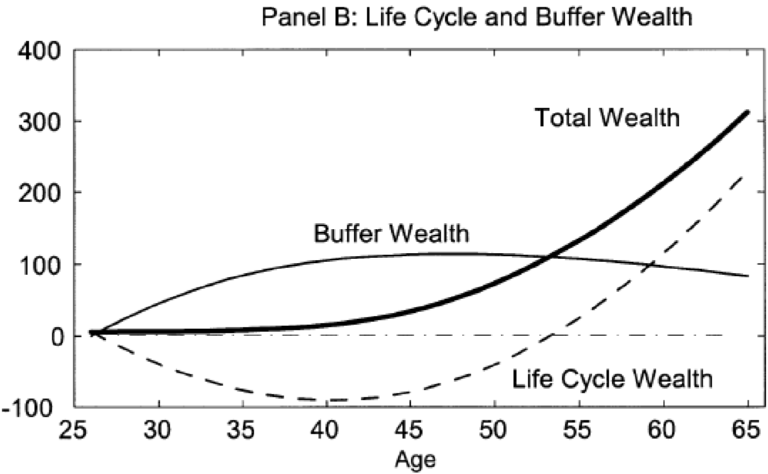


FIGURE 7.—The role of risk in saving and wealth accumulation.

Mathematical Programming with Equilibrium Constraints (MPEC)

- **Idea:** Do not solve the model, treat it as a constraint
- **Example:** Infinite horizon buffer-stock consumption model

$$\begin{aligned}\hat{\theta}, \hat{c}_1, \ldots, \hat{c}_\# &= \arg\max_{\theta, c_1, \ldots, c_\#} \mathcal{L}(\theta) \\ &\text{s.t.} \\ 0 &\leq c_j \leq m_j \\ 0 &\geq \mathcal{E}_j \\ 0 &= (m_j - c_j)\mathcal{E}_j\end{aligned}$$

where $\mathcal{E}_j$ is the j'th Euler-residual

$$\mathcal{E}_j \equiv \beta R \mathbb{E}_t[(G\psi_{t+1}c_{t+1}(\frac{1}{G\psi_{t+1}}Ra_i + \xi_{t+1}))^{-\rho}] - c_j^{-\rho}$$

and $c_{t+1}(\bullet)$ is interpolated using $c_1, c_2, \ldots, c_\#$

- **See Jørgensen (Economic Letters, 2013)**
- Intractable for life cycle models: Here EGM in a nested loop is much, much faster