

Sequence Space Jacobians

ECO2101 Lecture 8

Sebastian Dyrda, University of Toronto

30, March 2026

Introduction

Last lecture: Applications of Krusell–Smith with heterogeneous firms

- Khan and Thomas (2008): idiosyncratic shocks and non-convex investment
- Khan and Thomas (2013): credit shocks and the Great Recession
- Bloom, Floetotto, Jaimovich, Saporta-Eksten, Terry (2018): uncertainty shocks

This lecture: Auclert, Bardóczy, Rognlie, and Straub (2021)

- A general method for solving heterogeneous-agent models
- Key idea: write equilibrium as $H(\mathbf{K}, \mathbf{Z}) = 0$ in the **sequence space**
- Impulse responses from a single matrix: $d\mathbf{K} = G d\mathbf{Z}$
- Connection to the stochastic model via **certainty equivalence**

Next lecture: The Fake News Algorithm, DAGs, and applications to HANK

The Sequence-Space Jacobian

Model Description: Households

- Mass 1 of households maximize $\mathbb{E}[\sum \beta^t u(c_t)]$ with CRRA utility $u(c) = \frac{c^{1-\sigma}}{1-\sigma}$
- n_e idiosyncratic states; agents transition between states e and e' with probability $P(e, e')$. Stationary distribution: π
- Bellman equation for agent in state (e, k_-) at time t :

$$V_t(e, k_-) = \max_{c, k} u(c) + \beta \sum_{e'} V_{t+1}(e', k) P(e, e') \quad (2)$$

subject to:

$$\begin{aligned} c + k &= (1 + r_t)k_- + w_t e n \\ k &\geq 0 \end{aligned}$$

- Optimal policies: $c_t^*(e, k_-)$ and $k_t^*(e, k_-)$

Model Description: Distribution

- $D_t(e, K_-) \equiv \Pr(e_t = e, k_{t-1} \in K_-)$: measure of households in state e with capital in set K_-
- Law of motion for the distribution:

$$D_{t+1}(e', K) = \sum_e D_t(e, k_t^{*-1}(e, K)) P(e, e') \quad (3)$$

where $k_t^{*-1}(e, \cdot)$ denotes the inverse of $k_t^*(e, \cdot)$

- **Initial conditions:** Prior to $t = 0$ the economy is in steady state with constant (w_{ss}, r_{ss}) . Unique stationary distribution D_{ss} solving (3). We set $D_0 = D_{ss}$

Model Description: The Capital Function

- From (2): the policy $k_t^*(e, k_-)$ is a function of the *future* path $\{r_s, w_s\}_{s \geq t}$
- From (3) with $D_0 = D_{ss}$: the distribution $D_t(e, K)$ is a function of the *entire* path $\{r_s, w_s\}_{s \geq 0}$
- Aggregate household capital holdings are characterized by a **capital function**:

$$\mathcal{K}_t(\{r_s, w_s\}_{s \geq 0}) = \sum_e \int_{k_-} k_t^*(e, k_-) D_t(e, dk_-) \quad (4)$$

- The ability to reduce interactions between heterogeneous agents to functions like K_t , which map **aggregate sequences** into **aggregate sequences**, is key to the SSJ method

Model Description: Production and Market Clearing

- Standard firm first-order conditions:

$$r_t = \alpha Z_t \left(\frac{K_{t-1}}{N_t} \right)^{\alpha-1} - \delta \quad (5)$$

$$w_t = (1 - \alpha) Z_t \left(\frac{K_{t-1}}{N_t} \right)^{\alpha} \quad (6)$$

where $N_t = N_{ss} = \sum \pi e n$.

- Substituting (5)–(6) into (4), capital market clearing:

$$H_t(\mathbf{K}, \mathbf{Z}) \equiv \mathcal{K}_t \left(\left\{ \alpha Z_s \left(\frac{K_{s-1}}{N_{ss}} \right)^{\alpha-1} - \delta, (1 - \alpha) Z_s \left(\frac{K_{s-1}}{N_{ss}} \right)^{\alpha} \right\}_{s \geq 0} \right) - K_t = 0 \quad (7)$$

where $\mathbf{K} = (K_0, K_1, \dots)'$. Given K_{-1} and $\mathbf{Z} = (Z_0, Z_1, \dots)'$, eq. (7) pins down $\{K_t\}$

Impulse Responses

- Apply the implicit function theorem to (7). The linear impulse response of capital to a transitory technology shock $d\mathbf{Z} = (dZ_0, dZ_1, \dots)'$:

$$d\mathbf{K} = -\mathbf{H}_{\mathbf{K}}^{-1} \mathbf{H}_{\mathbf{Z}} d\mathbf{Z} \quad (8)$$

- **Local / first-order approximation:**

- We linearize $H(\mathbf{K}, \mathbf{Z}) = 0$ around steady state
- $d\mathbf{K}$ is the *best linear approximation* to the true response
- Valid for *small shocks*: ignores terms of order $(dZ)^2$

- **Why does $d\mathbf{K}$ pin down everything else?**

- Prices are functions of (K_{t-1}, Z_t) : $r_t = r(K_{t-1}, Z_t)$, $w_t = w(K_{t-1}, Z_t)$
- Output and consumption are functions of (K_{t-1}, Z_t) and policies
- Once $\{dK_t\}$ is known, all other variables follow by substitution

The Chain Rule

- Differentiating (7) w.r.t. K_s , the (t, s) entry of H_K :

$$[\mathbf{H_K}]_{t,s} = \frac{\partial \mathcal{K}_t}{\partial r_{s+1}} \frac{\partial r_{s+1}}{\partial \mathcal{K}_s} + \frac{\partial \mathcal{K}_t}{\partial w_{s+1}} \frac{\partial w_{s+1}}{\partial \mathcal{K}_s} - \mathbf{1}\{s = t\} \quad (9)$$

- A similar expression applies to H_Z
- The price derivatives are **analytical** from (5)–(6) at $(\mathbf{K}_{ss}, \mathbf{Z}_{ss})$. For example:

$$\frac{\partial r_{s+1}}{\partial \mathcal{K}_s} = \alpha(\alpha - 1)Z_{ss} \left(\frac{K_{ss}}{N_{ss}} \right)^{\alpha-2} \cdot \frac{1}{N_{ss}}$$

Key remaining challenge

To obtain $H_K^{-1} H_Z$ in (8), all we need are the Jacobians of the $\mathcal{K}$ function w.r.t. its two inputs r and w . How do we compute these? Later more on that.

Sequence-Space Jacobians: Rows, Columns, Dimensions

$$\mathbf{K} = (K_0, \dots, K_T)', \quad \mathbf{Z} = (Z_0, \dots, Z_T)', \quad \mathbf{H}(\mathbf{K}, \mathbf{Z}) = (H_0, \dots, H_T)'$$

Jacobian w.r.t. endogenous path $\mathbf{K}$

Jacobian w.r.t. exogenous path $\mathbf{Z}$

$$H_K = \begin{matrix} (T+1) \times (T+1) \\ \begin{bmatrix} H_0 & K_0 & K_1 & \cdots & K_T \\ \frac{\partial H_0}{\partial K_0} & \frac{\partial H_0}{\partial K_1} & \cdots & \frac{\partial H_0}{\partial K_T} \\ H_1 & \frac{\partial H_1}{\partial K_0} & \frac{\partial H_1}{\partial K_1} & \cdots & \frac{\partial H_1}{\partial K_T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_T & \frac{\partial H_T}{\partial K_0} & \frac{\partial H_T}{\partial K_1} & \cdots & \frac{\partial H_T}{\partial K_T} \end{bmatrix} \end{matrix} \quad H_Z = \begin{matrix} (T+1) \times (T+1) \\ \begin{bmatrix} H_0 & Z_0 & Z_1 & \cdots & Z_T \\ \frac{\partial H_0}{\partial Z_0} & \frac{\partial H_0}{\partial Z_1} & \cdots & \frac{\partial H_0}{\partial Z_T} \\ H_1 & \frac{\partial H_1}{\partial Z_0} & \frac{\partial H_1}{\partial Z_1} & \cdots & \frac{\partial H_1}{\partial Z_T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H_T & \frac{\partial H_T}{\partial Z_0} & \frac{\partial H_T}{\partial Z_1} & \cdots & \frac{\partial H_T}{\partial Z_T} \end{bmatrix} \end{matrix}$$

- **Rows:** equilibrium conditions at dates $t = 0, \dots, T$
- **Columns:** dated unknowns K_s or dated shocks Z_s , $s = 0, \dots, T$
- **One column** = effect of changing one date- s object on all equilibrium conditions

The Impulse Response Matrix $G \equiv -H_K^{-1} H_Z$

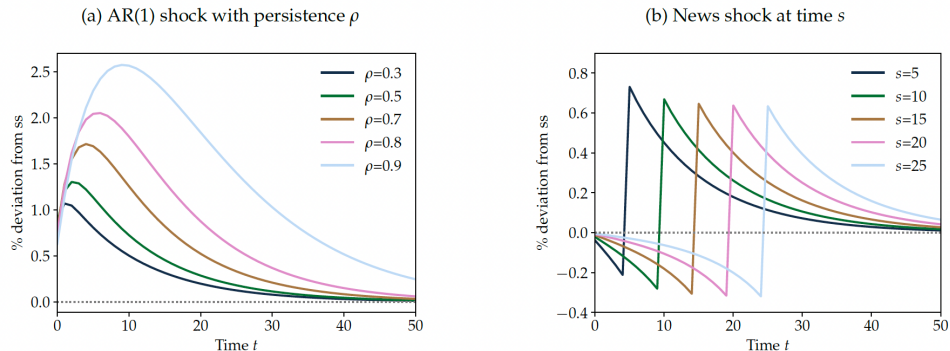
Given $d\mathbf{K} = -H_K^{-1} H_Z d\mathbf{Z}$, define $G \equiv -H_K^{-1} H_Z$:

$$G_{(T+1) \times (T+1)} = \begin{bmatrix} & Z_0 & Z_1 & \cdots & Z_T \\ K_0 & \frac{dK_0}{dZ_0} & \frac{dK_0}{dZ_1} & \cdots & \frac{dK_0}{dZ_T} \\ K_1 & \frac{dK_1}{dZ_0} & \frac{dK_1}{dZ_1} & \cdots & \frac{dK_1}{dZ_T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ K_T & \frac{dK_T}{dZ_0} & \frac{dK_T}{dZ_1} & \cdots & \frac{dK_T}{dZ_T} \end{bmatrix}$$

- **Column** s : impulse response of $\{K_t\}_{t=0}^T$ to a unit shock $dZ_s = 1$ (news shock at date s)
- **Row** t : sensitivity of K_t to shocks at every date
- **Any IRF**: $d\mathbf{K} = G d\mathbf{Z}$ — a single matrix-vector multiplication
- Example: AR(1) with persistence $\rho \Rightarrow d\mathbf{Z} = (1, \rho, \rho^2, \dots)'$ $\Rightarrow dK_t = \sum_s G_{t,s} \rho^s$

Impulse Responses of Capital to 1% TFP Shocks

Figure 1: Impulse responses of capital to 1% TFP shocks in the Krusell-Smith model



The same matrix $-H_K^{-1}H_Z$ is applied to all $d\mathbf{Z}$ vectors: once computed, additional IRFs are almost costless (single matrix-vector multiplication)

News Shock Interpretation

Panel (b): column s of $G = -H_K^{-1}H_Z$ = response of $\{K_t\}$ to a **one-time** 1% TFP increase at date s only

Panel (a): AR(1) TFP with persistence ρ means TFP is higher at every date:

$$d\mathbf{Z} = \underbrace{(1, \rho, \rho^2, \dots)'}_{\text{entire path known at date 0}}$$

Under perfect foresight, this is the *simultaneous* arrival of news about all future dates. The IRF is a weighted sum of the columns of G :

$$dK_t = \sum_{s=0}^T G_{t,s} \rho^s = \underbrace{G_{t,0}}_{\substack{\text{news:} \\ Z_0 \text{ up by } 1}} + \underbrace{\rho G_{t,1}}_{\substack{\text{news:} \\ Z_1 \text{ up by } \rho}} + \underbrace{\rho^2 G_{t,2}}_{\substack{\text{news:} \\ Z_2 \text{ up by } \rho^2}} + \dots$$

Given G , the underlying heterogeneity **no longer matters**: the Jacobians tell us everything we need to know, to first order, about the aggregate behavior of the model's heterogeneous agents.

From Perfect Foresight to Aggregate Uncertainty

How Does This Relate to the Stochastic KS Model?

We computed IRFs under **perfect foresight** over aggregates

But the Krusell–Smith model has **genuine aggregate uncertainty**:

- Z_t is stochastic (e.g. AR(1): $\log Z_t = \rho \log Z_{t-1} + \sigma \varepsilon_t$)
- Agents must *forecast* future prices
- The distribution D_t is a state variable — high-dimensional

Three approaches in the literature to deal with this:

1. **Krusell & Smith (1998)**: solve nonlinearly, approximate aggregation
2. **Boppart, Krusell & Mitman (2018)**: MIT shocks
3. **Reiter (2009)**: linearize in state space

We will see that, to first order, they all give the *same answer* as the SSJ approach — but at very different computational costs

Approach 1: Krusell & Smith (1998)

The problem:

- Full state of the economy at date t : (D_t, Z_t)
- D_t lives on a grid of n_g points (e.g. $n_g = 3,500$ or $250,000$)
- Agents need a forecasting rule: $r_t = r(D_t, Z_t)$, $w_t = w(D_t, Z_t)$
- Solving this exactly is intractable

The KS fix – approximate aggregation:

- Conjecture that agents only need to track $\bar{K}_t = \int k dD_t$ instead of D_t
- Forecasting rule: $\log \bar{K}_{t+1} = a_0 + a_1 \log \bar{K}_t + a_2 \log Z_t$
- Solve by simulation: draw shocks, simulate a panel, update (a_0, a_1, a_2) , iterate
- Verify: check that R^2 of the forecasting regression ≈ 1

Limitation: works because KS model happens to exhibit approximate aggregation. Not guaranteed in richer models (e.g. two-asset HANK)

Approach 2: Boppart, Krusell & Mitman (2018)

Key idea: use perfect-foresight transitions to learn about the stochastic model

MIT shock method:

1. Start the economy in steady state at $t = 0$
2. Hit it with a *one-time unexpected* shock to Z_0 (an “MIT shock”)
3. From $t = 0$ onward: agents have perfect foresight over the transition back
4. Solve for the nonlinear perfect-foresight path $\{K_t\}_{t=0}^T$

Connection to the stochastic model:

- For *small* shocks, this nonlinear transition $\approx$ the linear IRF of the stochastic model
- Why? Because to first order, the stochastic and perfect-foresight paths coincide

Limitation: BKM solve the nonlinear transition by iterating over guesses for $\{K_t\}$ — no general method for updating guesses, no guarantee of convergence

What BKM Showed Us

BKM's crucial insight: the linear IRFs form an **MA representation** of the stochastic model

If dK_s is the IRF at horizon s to a unit shock, then the stochastic process for capital is:

$$d\tilde{K}_t = \sum_{s=0}^{T-1} dK_s \varepsilon_{t-s}$$

From this $\text{MA}(T-1)$ representation:

- **Simulate:** draw $\{\varepsilon_t\}$, evaluate the sum
- **Second moments:** $\text{Cov}(d\tilde{K}_t, d\tilde{K}_{t+h}) = \sum_{s=0}^{T-1-h} (dK_s)(dK_{s+h})$

So if we can get the IRFs, we can fully characterize the stochastic model (to first order)

But BKM compute IRFs one at a time, nonlinearly, with an iterative solver

⇒ Slow, and hard to use for estimation (need to recompute for each parameter draw)

Approach 3: Reiter (2009)

Key idea: linearize the stochastic model directly

Setup:

- Write *all* equilibrium conditions – Bellman equation (2), distribution (3), and market clearing – as a system in (D_t, v_t, K_t, Z_t)
- Linearize around the steady state w.r.t. *aggregates*
- Keep full nonlinearity w.r.t. idiosyncratic shocks (the grid is exact)

Result: a linear rational expectations system in state space:

$$A \mathbb{E}_t[d\mathbf{X}_{t+1}] + B d\mathbf{X}_t + C d\mathbf{X}_{t-1} + E \varepsilon_t = 0$$

where $d\mathbf{X}_t$ stacks $(dv_t, dD_{t+1}, dK_t, dZ_t)$ – dimension $\approx 2n_g + 2$

Solve with standard methods (e.g. Schur / QZ decomposition)

This is exact (for the linearized problem) – no approximate aggregation needed

The Problem with Reiter

The system $A \mathbb{E}_t[d\mathbf{X}_{t+1}] + B d\mathbf{X}_t + C d\mathbf{X}_{t-1} + E \varepsilon_t = 0$ has dimension $\approx 2n_g + 2$

Schur decomposition costs $O(n_g^3)$:

- $n_g = 3,500$ (standard KS): feasible
- $n_g = 250,000$ (high-dimensional KS): **infeasible**
- $n_g = 10,500$ (two-asset HANK): requires model reduction

Model reduction (Ahn et al. 2018, Winberry 2018, Bayer & Luetticke 2020):

- Approximate D_t with fewer states (e.g. splines, moments)
- Reduces the system size, but introduces approximation error
- Accuracy depends on the application

$\Rightarrow$ Reiter gives the right answer, but **scales with the grid size n_g**

Certainty Equivalence: Definition

Certainty Equivalence (Simon, 1956; Theil, 1957)

Consider an optimizing agent facing uncertainty about future states. If the agent's decision rules are **linear** in the state variables, then the optimal decision under uncertainty is the same as the decision obtained by replacing all random variables with their conditional expectations.

Why does this matter here?

- The stochastic KS model: agents face aggregate uncertainty over Z_{t+1}
- First-order perturbation in aggregates $\Rightarrow$ all aggregate equilibrium conditions become *linear* in $(dK, dZ, \dots)$
- Certainty equivalence then applies: the linearized dynamics are identical whether agents face aggregate uncertainty or have perfect foresight

$\Rightarrow$ The perfect-foresight IRFs we computed **are** the first-order dynamics of the stochastic model

Certainty Equivalence: The Linearization Step

In the stochastic model, agents solve:

$$V_t(e, k_-) = \max_{c, k} u(c) + \beta \sum_{e'} \mathbb{E}_t[V_{t+1}(e', k)] P(e, e')$$

where $\mathbb{E}_t$ is over *aggregate* uncertainty Z_{t+1} (idiosyncratic risk handled by $\sum_{e'}$)

The *nonlinear* aggregate equilibrium conditions take the form:

$$\mathbb{E}_t[F(K_{t+1}, K_t, Z_{t+1}, Z_t)] = 0$$

First-order perturbation (Reiter, 2009): Taylor expand around steady state:

$$\mathbb{E}_t \left[\underbrace{\frac{\partial F}{\partial K_{t+1}}}_{\text{constant}} dK_{t+1} + \underbrace{\frac{\partial F}{\partial K_t}}_{\text{constant}} dK_t + \underbrace{\frac{\partial F}{\partial Z_{t+1}}}_{\text{constant}} dZ_{t+1} + \underbrace{\frac{\partial F}{\partial Z_t}}_{\text{constant}} dZ_t \right] = 0$$

Call this $\mathbb{E}_t[g(dK_{t+1}, dK_t, dZ_{t+1}, dZ_t)] = 0$

g is **linear by construction** — it is the first-order Taylor expansion, with all partial derivatives evaluated at steady state (constants)

Certainty Equivalence: From Stochastic to Deterministic

We have:

$$\mathbb{E}_t[g(dK_{t+1}, dK_t, dZ_{t+1}, dZ_t)] = 0$$

Since g is linear, $\mathbb{E}_t$ passes through (expectation of a linear function = linear function of expectations):

$$g(\mathbb{E}_t[dK_{t+1}], dK_t, \mathbb{E}_t[dZ_{t+1}], dZ_t) = 0$$

Under perfect foresight, $\mathbb{E}_t[dK_{t+1}] = dK_{t+1}$ and $\mathbb{E}_t[dZ_{t+1}] = dZ_{t+1}$:

$$g(dK_{t+1}, dK_t, dZ_{t+1}, dZ_t) = 0$$

Same equation. Certainty equivalence is not an assumption – it is a **consequence** of linearization

The logic:

1. Nonlinear stochastic conditions: $\mathbb{E}_t[F(\cdot)] = 0$
2. Linearize (Taylor expand) $\Rightarrow g$ is linear by construction
3. Linearity of $g \Rightarrow \mathbb{E}_t[g(\cdot)] = g(\mathbb{E}_t[\cdot]) \Rightarrow$ certainty equivalence
4. $\Rightarrow$ Perfect-foresight and stochastic systems give the **same first-order answer**

What Certainty Equivalence Buys Us

In the **stochastic** model:

- Agents must forecast future prices: $r_{t+1} = r(D_{t+1}, Z_{t+1})$
- Need to track the distribution D_t as a state variable
- This is what makes KS hard: D_t is n_g -dimensional

After **linearization + certainty equivalence**:

- The problem is deterministic: agents know the entire future path of prices $\{r_s, w_s\}_{s \geq t}$
- No need to forecast $\Rightarrow$ no need to track D_t as a state variable
- **Calendar time t is a sufficient statistic** for the aggregate environment

$\Rightarrow$ We can work directly with **sequences** $\{K_t, Z_t, r_t, w_t\}_{t=0}^T$ instead of a recursive law of motion over a high-dimensional state space

The SSJ Approach: Best of All Worlds

Certainty equivalence implies: SSJ, Reiter, and BKM (for small shocks) all give the **same first-order answer**

But the SSJ system $H(\mathbf{K}, \mathbf{Z}) = 0$ has dimension $n_u \times T$:

- n_u : number of *unknowns* (e.g. 1 for KS, 3 for two-asset HANK)
- T : truncation horizon (e.g. 300)
- **Independent of n_g**

All the heterogeneity is absorbed into the Jacobians of the K function, computed *once*

Once we have $G = -H_K^{-1} H_Z$:

- Simulation, moments, likelihood via the MA representation (BKM insight)
- Recomputing IRFs for new shock parameters: just matrix multiplication
- Bayesian estimation becomes feasible even for two-asset HANK

Comparing the Four Approaches

	KS (1998)	BKM (2018)	Reiter (2009)	SSJ (2021)
Linearizes?	No	No	Yes	Yes
Aggregation	Approximate	Exact	Exact	Exact
System size	Small	T	$\sim 2n_g$	$n_u \times T$
Depends on n_g ?	No*	Yes**	Yes	No
IRF computation	Simulation	Iterative	Schur	Fake news
Estimation	Hard	Hard	Feasible***	Fast

* Small system, but requires approximate aggregation (may be inaccurate)

** Each nonlinear transition requires backward/forward iteration on the full grid

*** Only if n_g is small enough for Schur decomposition

Next lecture: The key computational challenge — how to efficiently compute the Jacobians of the K function. This is the **Fake News Algorithm** (Section 3)

This Lecture: Summary

Section 2 of Auclert et al. (2021):

- Krusell–Smith model written as $H(\mathbf{K}, \mathbf{Z}) = 0$ in the sequence space
- Impulse responses via the IFT: $d\mathbf{K} = G d\mathbf{Z}$ where $G = -H_K^{-1} H_Z$
- Columns of G = news shock IRFs; any IRF is a matrix-vector product

Connecting to the stochastic model:

- KS (1998): nonlinear, approximate aggregation
- BKM (2018): MIT shocks $\Rightarrow$ IRFs $\Rightarrow$ MA representation
- Reiter (2009): linearize in state space, exact but scales with n_g
- Certainty equivalence: all approaches give the same first-order answer
- SSJ: system size independent of $n_g \Rightarrow$ scales to rich models

Open question: How do we actually compute the Jacobians of the $\mathcal{K}$ function?

Next Lecture: Computing Jacobians and Applications

Section 3: The Fake News Algorithm

- General representation of HA problems: eqs. (10)–(12)
- Policy symmetry (Lemma 1), fake news matrix (Lemma 2)
- Expectation vectors, Proposition 1: from F to $\mathcal{J}$
- The four steps of the algorithm; efficiency gains

Section 4: General Equilibrium via DAGs

- Composing Jacobians from different blocks
- Directed acyclic graphs (DAGs) for model representation
- Forward accumulation of Jacobians along the DAG

Applications:

- One-asset HANK: monetary policy, fiscal policy
- Two-asset HANK: liquid vs illiquid, portfolio adjustment
- Bayesian estimation of HANK models on US data

Readings

- Auclert, Bardóczy, Rognlie, and Straub (2021), “Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models,” *Econometrica*, 89(5), 2375–2408
`doi:10.3982/ECTA17434`
 - This lecture: Sections 2 and introduction to Section 3
 - Next lecture: Sections 3–4, selected results from 5–6
- Krusell and Smith (1998), “Income and Wealth Heterogeneity in the Macroeconomy,” *JPE*, 106(5), 867–896, `doi:10.1086/250034`
- Boppart, Krusell, and Mitman (2018), “Exploiting MIT Shocks in Heterogeneous-Agent Economies,” *JEDC*, 89, 68–92, `doi:10.1016/j.jedc.2018.01.002`
- Reiter (2009), “Solving Heterogeneous-Agent Models by Projection and Perturbation,” *JEDC*, 33(3), 649–665, `doi:10.1016/j.jedc.2008.08.010`

Code: <https://github.com/shade-econ/sequence-jacobian>