Exploiting MIT Shocks in Heterogenous-Agent Economies

Boppart, Krusell and Mitman

R.S Chhina

November 17, 2019

Before we begin

- I have only a few slides on the actual BKM paper. I rather try to present an overview of the entire literature.
- Most of my knowledge on these methods comes from the few weeks I spent on them during my dissertation, so might say some stupid things, please correct!
- The paper and code is in discrete time but I use continuous time.

The Model

- Standard Aiyagari (1994) economy with aggregate shocks.

Household's problem

$$\max_{\{c_t\},t>0} \mathbb{E} \int_0^\infty e^{-\rho t} u(c_t) dt \tag{1}$$

$$\dot{a}_t = w_t z_t + r_t a_t - c_t \tag{2}$$

$$a_t \geq \underline{a} \tag{3}$$

$z_t \in \{z_l, z_h\}$, Poisson with intensities λ_l and λ_h. (Could be a diffusion process)

Firms

Representative firm production function

$$Y_t = e^{Z_t} K_t^{\alpha} N_t^{1-\alpha} \qquad dZ_t = -\theta Z_t dt + \eta dB_t \tag{4}$$

First Order Conditions

$$w_t = (1-\alpha)\frac{Y_t}{N_t} \tag{5}$$

$$r_t = \alpha \frac{Y_t}{K_t} - \delta \tag{6}$$

Market Clearing Conditions

$$K_t = \int a g_t(a,z) da dz \tag{7}$$

$$N_t = \int z g_t(a,z) da dz \equiv 1 \tag{8}$$

Stationary Equilibrium without aggregate uncertainty

Hamilton Jacobi Bellman (HJB)

$$\rho v(a,z) = \max_c u(c) + \frac{\partial v(a,z)}{\partial a}(wz + ra - c) + \lambda_z (v(a,z') - v(a,z)) \quad (9)$$

Kolmogrov Forward Eqn. (KFE)

$$0 = -\frac{\partial s(a,z) g(a,z)}{\partial a} - \lambda_z g(a,z) + \lambda_{z'} g(a,z') \quad (10)$$

Borrowing Condition

$$v'(\underline{a}, z) \geq u'(wz + r\underline{a}) \quad (11)$$

and the First Order and the Market Clearing conditions.

Numerical Method (for partial equilibrium)

Discretize the HJB equation; Use an upwind scheme to calculate derivatives

$$
\begin{aligned}
\frac{v_{i,j}^{n+1} - v_{i,j}^{n}}{\Delta} + \rho v_{i,j}^{n+1} = u(c_{i,j}^{n}) \\
+(v_{i,j,F}^{n+1})'[z_j + ra_i - c_{i,j,F}^{n}]^{+} + (v_{i,j,B}^{n+1})'[z_j + ra_i - c_{i,j,B}^{n}]^{-} \\
+\lambda_j[v_{i,-j}^{n+1} - v_{i,j}^{n+1}]
\end{aligned}
\tag{12}
$$

Iterate on this system of linear equations, until convergence.

$$
\frac{1}{\Delta}(v^{n+1} - v^{n}) + \rho v^{n+1} = u^{n} + \mathbf{A}^{n} v^{n+1} \tag{13}
$$

Solve KFE for free:

$$
\mathbf{A}^{\mathbf{T}} g = 0 \tag{14}
$$

Transition Dynamics

Transition path from a arbitrary initial condition or after a "MIT" shock. Numerical method goes back to Auerbach and Koltikoff (1987)

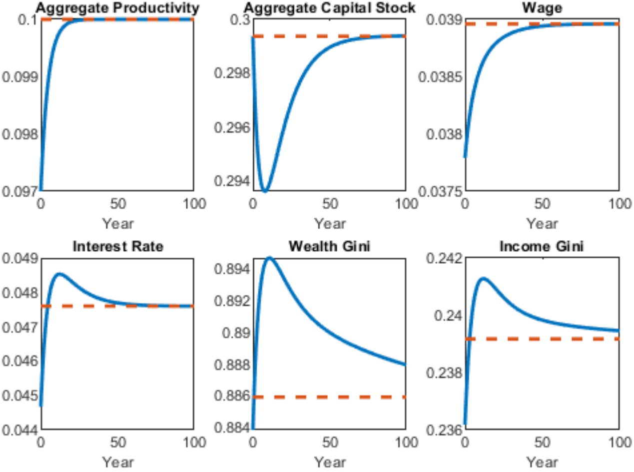
- What is a "MIT" shock?
- The economy eventually reaches the steady state at time T.
- Guess the path of K_{t+s}, s=0,1,2...T. Use FOC's to calculate w_t and r_t.

The HJB is now time dependent. Discretize in time(!!) now addition to wealth and write $v_{i,j}^n = v_j(a_i, t^n)$.

- Now each n has the interpretation of time step instead of an iteration on the stationary value function.
- Solve HJB backwards from the steady state and KFE "forwards".
- update the guess on the transition path of K up till convergence.

MIT shock

$\rho = 0.8$. Code by Ben Moll



Notice the distributional dynamics. For ex. could the proportion of wealthy HtM consumers in more complex models like last week's kaplan and Violante (2014) model.

Equilibrium with aggregate uncertainty

Now we solve the full model with aggregate shocks.
BIG PROBLEM: Whole distribution is a state variable.

$$w(g, Z) = (1-\alpha)e^Z K(g)^\alpha \qquad r(g, Z) = \alpha e^Z K(g)^{\alpha-1} - \delta \tag{15}$$

$$K(g) = \int a g(a, z) da dz \tag{16}$$

And just to show how big a mess an infinite dimensional HJB is

$$\begin{aligned}\rho V(a, z, g, Z) = \max_c u(c) &+ \partial_a V(a, z, g, Z)[w(g, Z)z + r(g, Z)a - c] \\ &+ \lambda_z [V(a, z', g, Z) - V(a, z, g, Z)] \\ &+ \partial_z V(a, z, g, Z)(-\nu Z) + \frac{1}{2}\partial_{ZZ} V(a, z, g, Z)\sigma^2 \\ &+ \int \frac{\delta V(a, z, g, Z)}{\delta g(a, z)} T[g, Z](a, z) da dz\end{aligned} \tag{17}$$

$$T[g, Z](a, z) = -\partial_a[s(a, z, g, Z)g(a, z)] - \lambda_z g(a, z) + \lambda_{z'} g(a, z') \tag{18}$$

g(a,z) is the whole distribution, an infinite dimensional object. Computationally highly infeasible.

Solution 1 - Approximate Aggregation - Krussel-Smith

partial information, bounded rationality etc

Key Idea: "Perhaps by using a little bit information about g(a,z), households do almost as well as by using all the information in the distribution of g(a,z) when predicting future prices" - Victor Rios-Rull (1997)

- Solution by simulation.
- Approximate aggregation: Instead of the whole distribution, just look at the mean (which is also the average capital stock) i.e. $K = \int ag(a,z)dadz$.
- And instead of using the KF operator the households have a "Perceived Law of Motion" (PLM) i.e. $\dot{K} = h(K, Z)$ (in discrete time $K' = \alpha + \beta K$ for KS).
- A differential game with a very large number of players. Agents optimize given the state of the economy. Given the agents controls and "common noise" the state of the economy evolves. K & S argue that this evolution just linearly depends upon the current aggregate.

Equilibrium

Given the PLM, households HJB now becomes
I have written this equation similar to Fernández-Villaverde, J., Hurtado, S., Nuno, G. (2019) (but could be wrong)

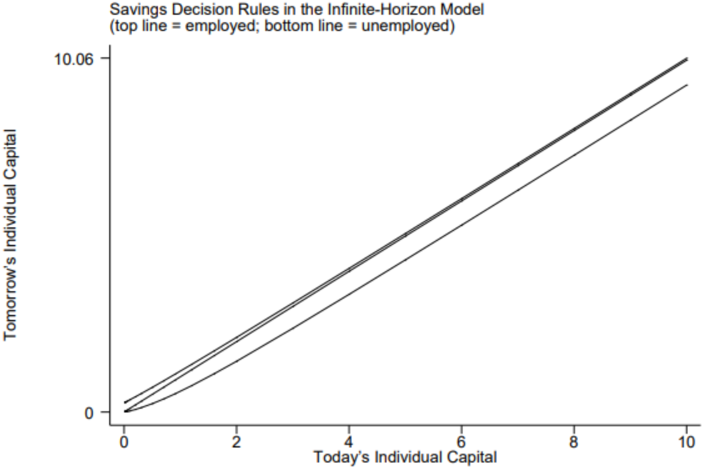
$$\begin{aligned}\rho V(a,z,K,Z) = \max_c u(c) + \partial_a V(a,z,K,Z)[w(K,Z)z + r(K,Z)a - c] \\ +\lambda_z[V(a,z',K,Z) - V(a,z,K,Z)] \\ +\partial_z V(a,z,K,Z)(-\nu Z) + \frac{1}{2}\partial_{ZZ} V(a,z,K,Z)\sigma^2 \\ +h(K,Z)\frac{\partial V(a,z,K,Z)}{\partial K}\end{aligned} \tag{19}$$

Algorithm

1. Guess the perceived law of motion h(K,Z)
2. Solve the household HJB using the finite differences method.
3. Simulate the model using this policy function for a cross section of individuals.
4. Use regression/universal non linear approximator/any other machine learning technique to update the guess.
5. Iterate till convergence

In Step 4: KS use regression to update their linear PLM; Fernandez-Villaverde (2018) use universal non linear approximator (Neural Network), Duarte (2018) provides an open source library with a number of general machine learning tools.

Why does Krusell-Smith works?



Why does Krusell-Smith works?

- Both in steady state and in aggregate uncertainty, decision rules are almost linear in individual wealth.
- Except for the very poor MPCs are identical.
- "very poor" regions is small and due to thin tail distribution carries little wealth, not affecting the aggregate.
- Moreover large re-distributions involving this population would be required, which seldom happens.

What causes these features?

- Although markets are incomplete, precautionary savings act as insurance against income shocks.
- Wealthier the agent, less important are these fluctuations to her.
- Hence, the MPCs are rather determined by permanent income considerations. Thus making the savings propensity independent of wealth for high enough wealth levels.
- Moreover, with long enough horizons, agents ex-ante save enough to avoid being in the lower region and distribution moves right endogenously.

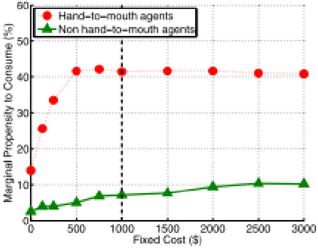
Does it always works?

"The boundaries of the applicability of approximate aggregation are far from known; **so far, we know of no quantitatively convincing models with large departures from aggregation** , but our imagination is admittedly limited and we foresee important examples of such phenomena to be discovered in future research." - Krussell and Smith (2006)

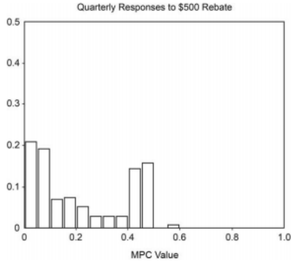
Have we found some now?

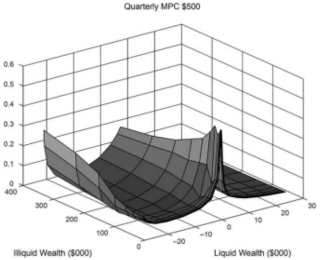
- Kaplan and Violante (2014) (discussed last week)
- BKM discuss - Ayigari(1994) with with valued leisure and two kinds of aggregate shocks - an input neutral technology shock and an investment specific technology shock.
- Demand Externality (consumption rises, so does TFP) in BKM inspired by Bai, Rios-Rull and Storesletten (2016)

Does it always works?



(b) Average marginal prop. to consume





Solution 2 - Linearization

Pioneered by Reiter(2009,2010), easily implementable toolbox available from Ahn et al. (2017) (Moll's website) **Notation**

$$\rho v_t(a,z) = \max_c u(c) + \partial_a v_t(a,z)(w_t z - r_t a - c) \\ + \lambda_z(v_t(a,z') - v_t(a,z)) + \frac{1}{dt} E_t[dv_t(a,z)] \tag{20}$$

$$\partial_t g_t(a,z) = -\partial_a[s_t(a,z)g_t(a,z)] - \lambda_z g_t(a,z) + \lambda_{z'} g_t(a,z') \tag{21}$$

$$dZ_t = -nuZ_t dt + \sigma dW_t \tag{22}$$

+ FOCs and Market clearing conditions.

Linearization

Fully Non-linear w.r.t individual shocks. Linear approximation to aggregate uncertainity. Three Steps:

- Compute the Stationary Steady State without aggregate uncertainty.
- Calculate First order Taylor expansions around the steady state (as in the standard RBC literature).
- Solve linear stochastic differential equations.

Linearization

- Discretized system with aggregate shocks

$$\begin{aligned}
\rho \mathbf{v}_t &= \mathbf{u}(\mathbf{v}_t) + \mathbf{A}(\mathbf{v}_t; \mathbf{p}_t)\, \mathbf{v}_t + \frac{1}{dt} \mathbb{E}_t d\mathbf{v}_t \\
\frac{d\mathbf{g}_t}{dt} &= \mathbf{A}(\mathbf{v}_t; \mathbf{p}_t)^{\mathsf{T}} \mathbf{g}_t \\
\mathbf{p}_t &= \mathbf{F}(\mathbf{g}_t; Z_t) \\
dZ_t &= -\nu Z_t dt + \sigma dW_t
\end{aligned}$$

- Linearize using automatic differentiation (code: `@myAD`)

$$\mathbb{E}_t \begin{bmatrix} d\widehat{\mathbf{v}}_t \\ d\widehat{\mathbf{g}}_t \\ \mathbf{0} \\ dZ_t \end{bmatrix} = \underbrace{\begin{bmatrix} \mathbf{B}_{vv} & \mathbf{0} & \mathbf{B}_{vp} & \mathbf{0} \\ \mathbf{B}_{gv} & \mathbf{B}_{gg} & \mathbf{B}_{gp} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{pg} & \mathbf{I} & \mathbf{B}_{pZ} \\ 0 & 0 & 0 & -\nu \end{bmatrix}}_{\mathsf{B}} \begin{bmatrix} \widehat{\mathbf{v}}_t \\ \widehat{\mathbf{g}}_t \\ \widehat{\mathbf{p}}_t \\ Z_t \end{bmatrix} dt$$

Linearization

- Size of the system manageable in simple applications. It's a system of 2N+3 equations in 2N+3 variables. N=2I; I is the number of grid points.
- Derivatives are calculated using Automatic Differentiation. It uses the fact that computer stores functions as a combination of elementary functions such as addition, multiplication, exponential etc. and differentiates by continuously applying the chain rule.
- Solve the system using Schur decomposition provided Blanchard-Kahn conditions hold etc etc

What does it looses?

- The solution features certainty equivalence to **aggregate shocks**. Standard deviation of Z, σ does not enters households decision rules.
- This means that linearization captures the effect of aggregate uncertainty only to the extent it affects the idiosyncratic shocks.
- **So not suitable for asset pricing models where aggregate uncertainty directly affects individual decision rules.**

Larger Models

- The method worked perfectly fine in smaller models like Krusell Smith but for a model like Kaplan and Violante (2014), the number of SDEs become prohibitively large.
- So, they use the standard model reduction techniques from engineering using projection methods.

  Jesús Fernández-Villaverde discusses some of them in his "Computational Methods for Economists" course.
- Carry out both Distribution and Value function reduction.
- For practical purposes just use the numerical toolbox they provide.

Results

The discretization of state space (a,b,z) contains N=60,000 points.

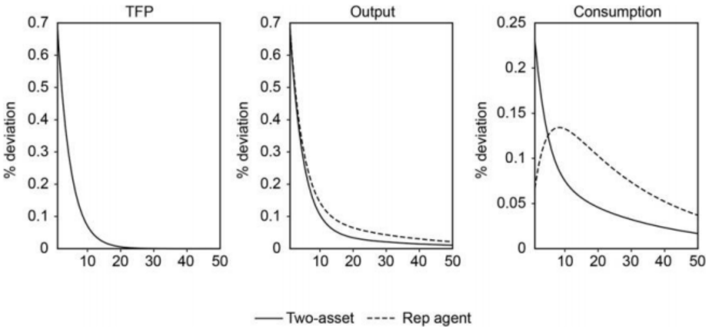


Fig. 9. Aggregate impulse responses to aggregate productivity shock Z_t

Figure: Kaplan and Violante with aggregate productivity shocks.

Finally! Boppart-Krusell-Mitman

What do we know up till now?

- How to calculate the response of aggregate variable x, $\{x_t\}_{t=0}^{T}$ to an aggregate TFP "MIT" shock.
- Well performing linearization methods can be designed using sophisticated techniques.

BUT if linearization works, then aggregate uncertainity is just a scaled additive combination of MIT shocks which we already know!!!!!

Or simply $x_0\epsilon_{z,T} + x_1\epsilon_{z,T-1} + x_2\epsilon_{z,T-2} + ...$ where $\epsilon_{z,t}$ is the innovation to z in period t.

BKM

Three key insights

- Dimensionality reduction is by using the **Sequence Form** instead of the recursive form.
- **Linearity** (already discussed)
- **Certainity Equivalance**: Assuming this allows to linearly add the MIT shock responses. As, the aggregate uncertainty has no affect on optimal policy rules. Note that it is not a drawback compared to linearization methods because they precisely lead to linear decision rules in which aggregate shocks have no effect.
- More than one shock? Compute two MIT responses and just add.

Michael Reiter's Comments on the Paper

- Is this method simpler than recursive linearized solution?
  Answer: **Hell Yes!**
  One downside is that it requires to calculate the transition on the perfect foresight path which doesn't have properly defined convergence properties.
- What is lost in the BKM method, compared to the recursive linear solution?
  Linearization method clearly tells if the model has a solution by checking Blanchard/Kahn conditions. Not sure how BKM performs if they are not satisfied.
- Better than linear? Well, see next slide.
  But one can check if the model has linear properties by looking at IRFs to shocks of different sizes, say 0.01σ and 2σ. If they are similar, model mut be linear to aggregate shocks?

Better than linear?

BKM consider a one standard deviation shock, which not small, so it might be linear superposition of a non linear response.

Consider a simple non linear model: $x_t = ax_{t-1} + bx_{t-1}^2 + z_t$

Table 1
Impulse responses, quadratic model.

	1	2	3
IR shock 1	$\bar{z}$	$a\bar{z} + b\bar{z}^2$	$a(a\bar{z} + b\bar{z}^2) + ba^2\bar{z}^2$
IR shock 2	0	$\bar{z}$	$a\bar{z} + b\bar{z}^2$
Linear superpos.	$\bar{z}$	$(a+1)\bar{z} + b\bar{z}^2$	$a((a+1)\bar{z} + b\bar{z}^2) + b(a^2+1)\bar{z}^2$
Exact response	$\bar{z}$	$(a+1)\bar{z} + b\bar{z}^2$	$a((a+1)\bar{z} + b\bar{z}^2) + b(a+1)^2\bar{z}^2$

Superposition of individual responses misses the quadratic term on the individual response.

Proposed Generalization

- Divide the support of the shock z into n intervals and for each interval calculate the separate Certainty Equivalent Impulse Response Function

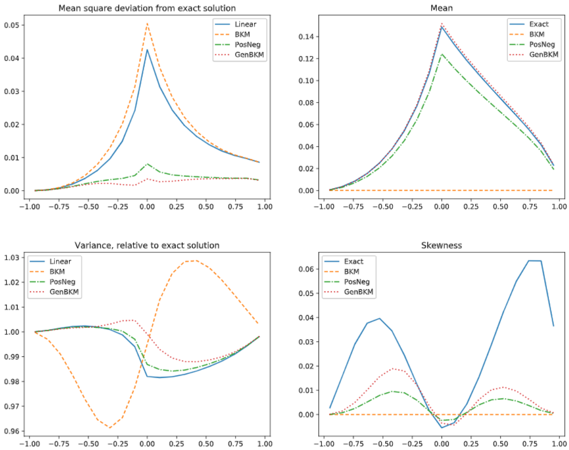


Fig. 1. Accuracy in nonlinear model.

Michael Reiter's Comments on the Paper

BKM is particularly useful in models with:

- The decision problem of the agent is complicated because it has several states, it may be non-convex, and it may have occasionally binding constraints. As a consequence, it must be solved on a discrete grid with many states. To give a number, say in the order of 10^6 states.
- Existence and uniqueness of stable solutions can be taken as given.
- The effect of uncertainty on behavior is known to be small, or in any case is not the focus of the analysis.
- The model is driven by only a few aggregate shocks.
  BUT uncertainty could shift the policy functions if certainty equivalence doesn't holds and linearity tests won't tell us this fact. Shortcoming of all these procesures.

Thanks