Lecture Notes on Macroeconomic Analysis with Machine Learning and Big Data

Weinan E and Yucheng Yang

2019 Summer @ Peking University

Please address any feedbacks or questions to yuchengy@princeton.edu

Macroeconomic Analysis with Machine Learning and Big Data

Lecture 1: Introduction

Weinan E   Yucheng Yang

July 2, 2019

Outline

Motivation of This Course

Big Data: New Way to Learn Economics?

The radical plan to change how Harvard teaches economics

Raj Chetty has an idea for introducing students to econ that could transform the field — and society.

By Dylan Matthews | @dylanmatt | dylan@vox.com | Updated May 22, 2019, 8:07am EDT

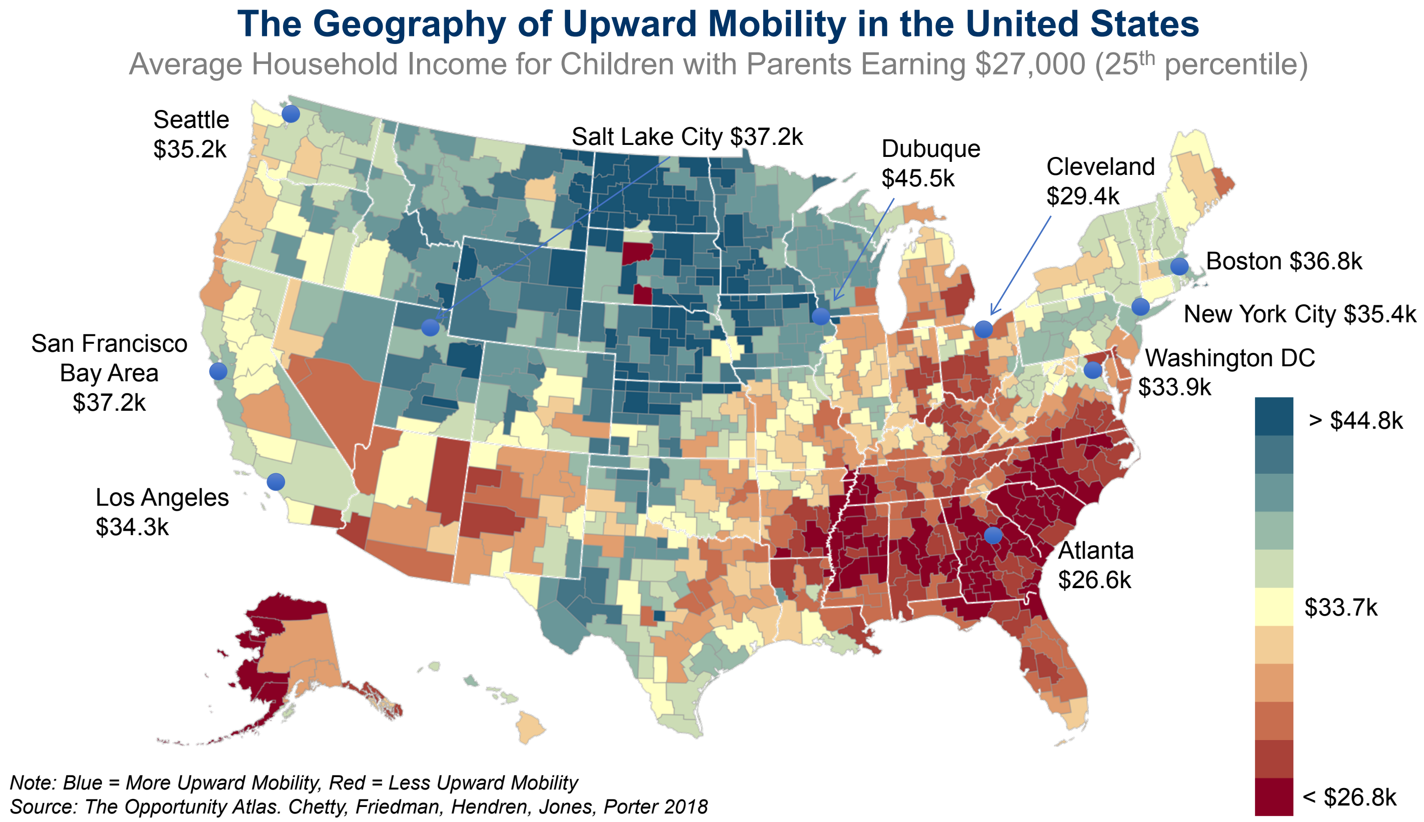
SHARE

The Highlight BY Vox

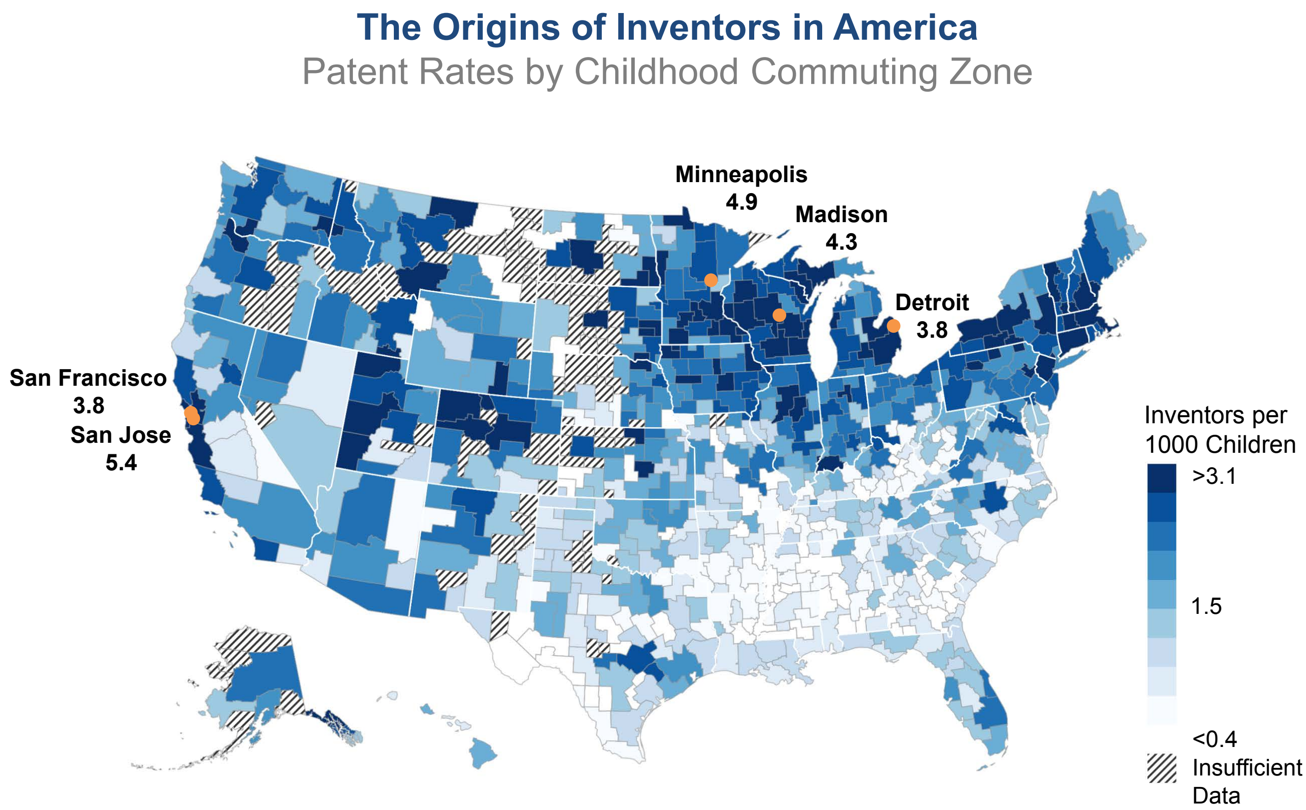
But another Harvard economist has a different idea of how to introduce students to economics.

Raj Chetty, a prominent faculty member whom Harvard recently poached back from Stanford, this spring unveiled **"Economics 1152: Using Big Data to Solve Economic and Social Problems."** Taught with the help of lecturer Greg Bruich, the class garnered 375 students, including 363 undergrads, in its first term. That's still behind the 461 in Ec 10 — but not by much.

Figure: Raj Chetty

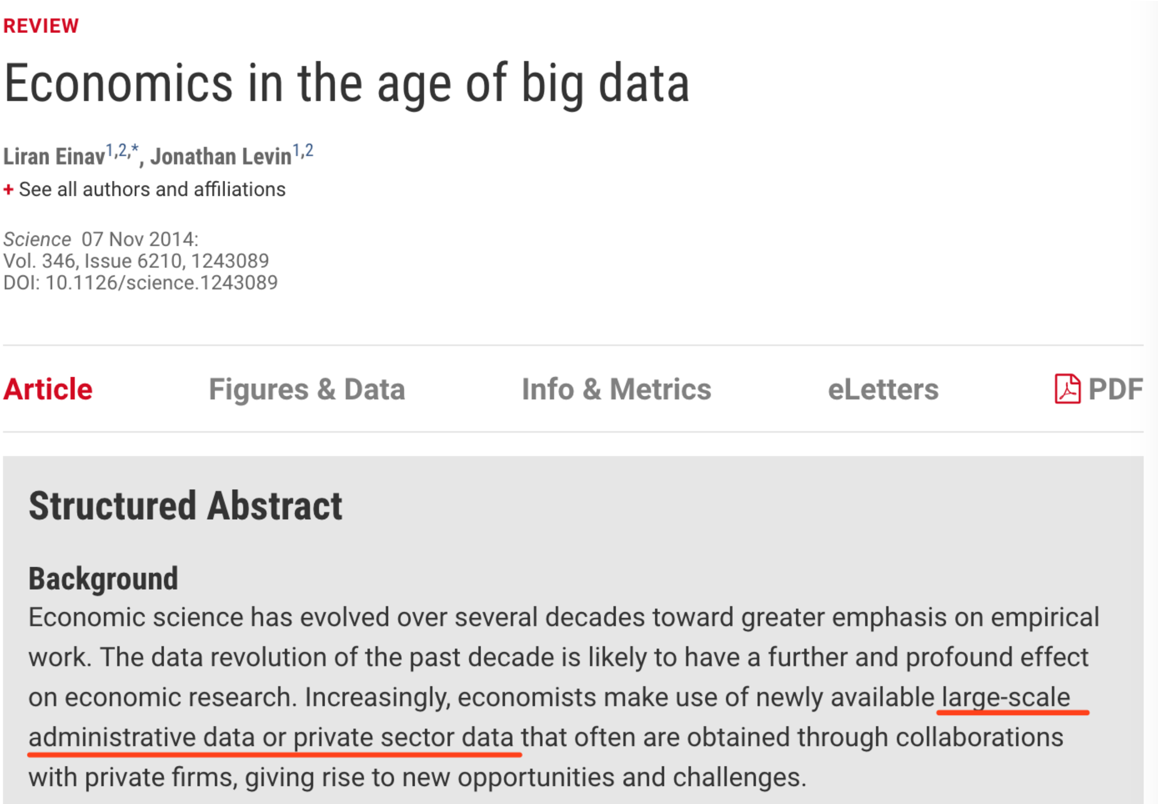


https://opportunityinsights.org/wp-content/uploads/2019/05/Lecture-1-intro-and-mobility.pdf
https://www.opportunityatlas.org/



https://opportunityinsights.org/wp-content/uploads/2019/05/Lecture-5-innovation.pdf

Big Data and Machine Learning: New Way to Study Economics?



REVIEW

Economics in the age of big data

Liran Einav[1,2,*], Jonathan Levin[1,2]
+ See all authors and affiliations

Science 07 Nov 2014:
Vol. 346, Issue 6210, 1243089
DOI: 10.1126/science.1243089

Article | Figures & Data | Info & Metrics | eLetters | PDF

Structured Abstract

Background
Economic science has evolved over several decades toward greater emphasis on empirical work. The data revolution of the past decade is likely to have a further and profound effect on economic research. Increasingly, economists make use of newly available large-scale administrative data or private sector data that often are obtained through collaborations with private firms, giving rise to new opportunities and challenges.

Figure: *Science* Review on the Economics in the Age of Big Data

Big Data and Machine Learning: New Way to Study Economics?

Economist.com

Figure: Upward Trend is persistent after 2016

Big Data and ML in Prominent Conferences for Macro

Figure: One Session in AEA 2019 Annual Conference

Big Data and ML in Prominent Conferences for Macro

Figure: One Session in AEA 2019 Annual Conference

Big Data and ML in Prominent Conferences for Macro

SI 2019 Micro Data and Macro Models

Erik Hurst, Greg Kaplan, and Giovanni L. Violante, Organizers
July 15-18, 2019
Longfellow Room
Royal Sonesta Hotel
40 Edwin H. Land Blvd.
Cambridge, MA

Conference Code of Conduct
Summer Institute 2019 master schedule

Monday, July 15

8:30 am	Coffee & Pastries
9:00 am	Rohan Kekre, University of Chicago Moritz Lenel, Princeton University *Redistribution, Risk Premia, and the Macroeconomy*
9:45 am	Laura Liu, Indiana University Mikkel Plagborg-Møller, Princeton University *Full-Information Estimation of Heterogeneous Agent Models Using Macro and Micro Data*
10:30 am	Break
10:45 am	Sushant Acharya, Federal Reserve Bank of New York Edouard Challe, Ecole Polytechnique Keshav Dogra, Federal Reserve Bank of New York *Optimal Monetary Policy in HANK Economies*

http://papers.nber.org/sched/SI19EFMM

Big Data and ML in Prominent Conferences for Macro

CRIW Conference: Big Data for 21st Century Economic Statistics

Katharine G. Abraham, Ron S. Jarmin, Brian Moyer, and Matthew D. Shapiro, Organizers
March 15-16, 2019
Supported by the Alfred P. Sloan Foundation

Hyatt Regency Bethesda
Cabinet/Judiciary Room
One Bethesda Metro Center
Bethesda, MD

Conference Code of Conduct

Friday, March 15

8:30 am	Continental Breakfast
	Session 1, Chair: Katharine Abraham, University of Maryland and NBER
9:00 am	Welcome from Conference Organizers and Opening Remarks from James Poterba, Massachusetts Institute of Technology and NBER
9:10 am	Gabriel Ehrlich, University of Michigan John C. Haltiwanger, University of Maryland and NBER Ron S. Jarmin, Bureau of the Census David Johnson, University of Michigan Matthew D. Shapiro, University of Michigan and NBER *Re-Engineering Key National Economic Indicators*

https://conference.nber.org/conferences/2019/CRIWs19/summary.html

Book on "Big Data for 21st Century Economic Statistics"

Table of Contents

https://www.nber.org/books/abra-7

Big Data and ML for Policy Makers

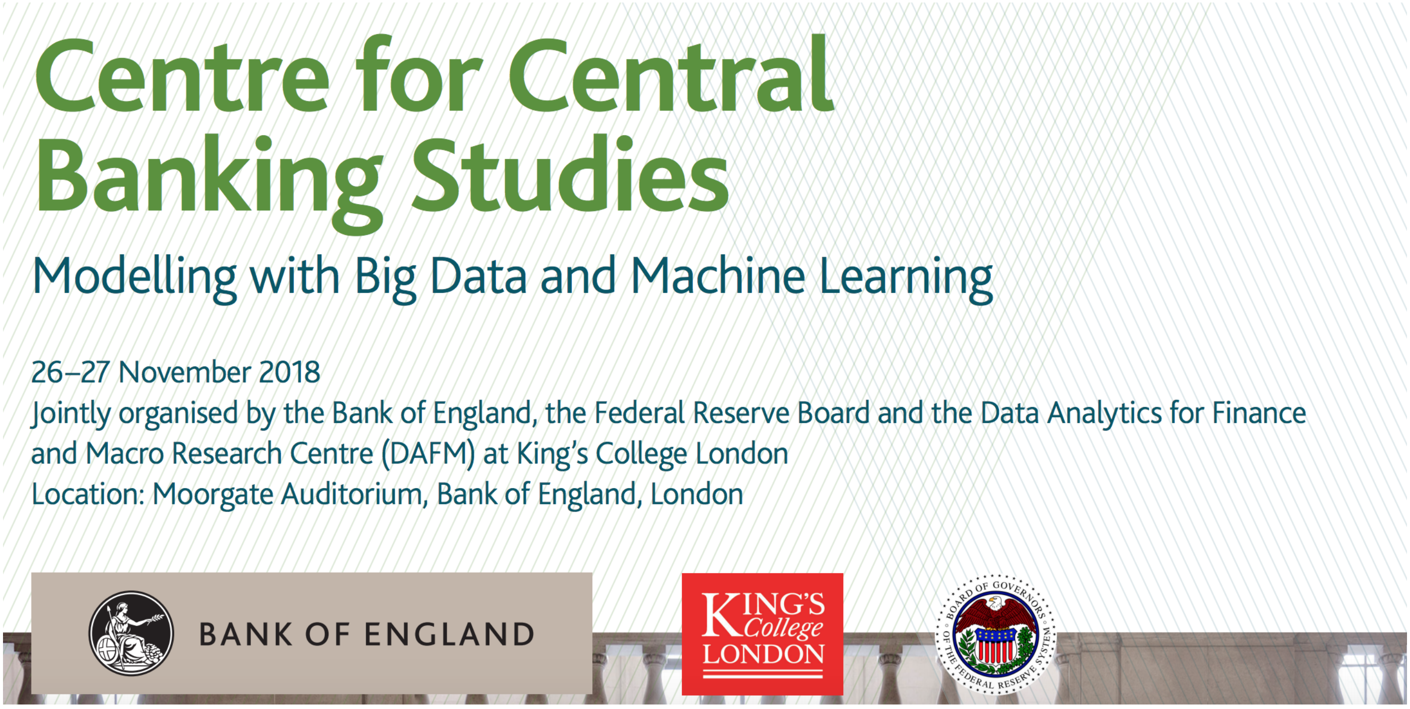
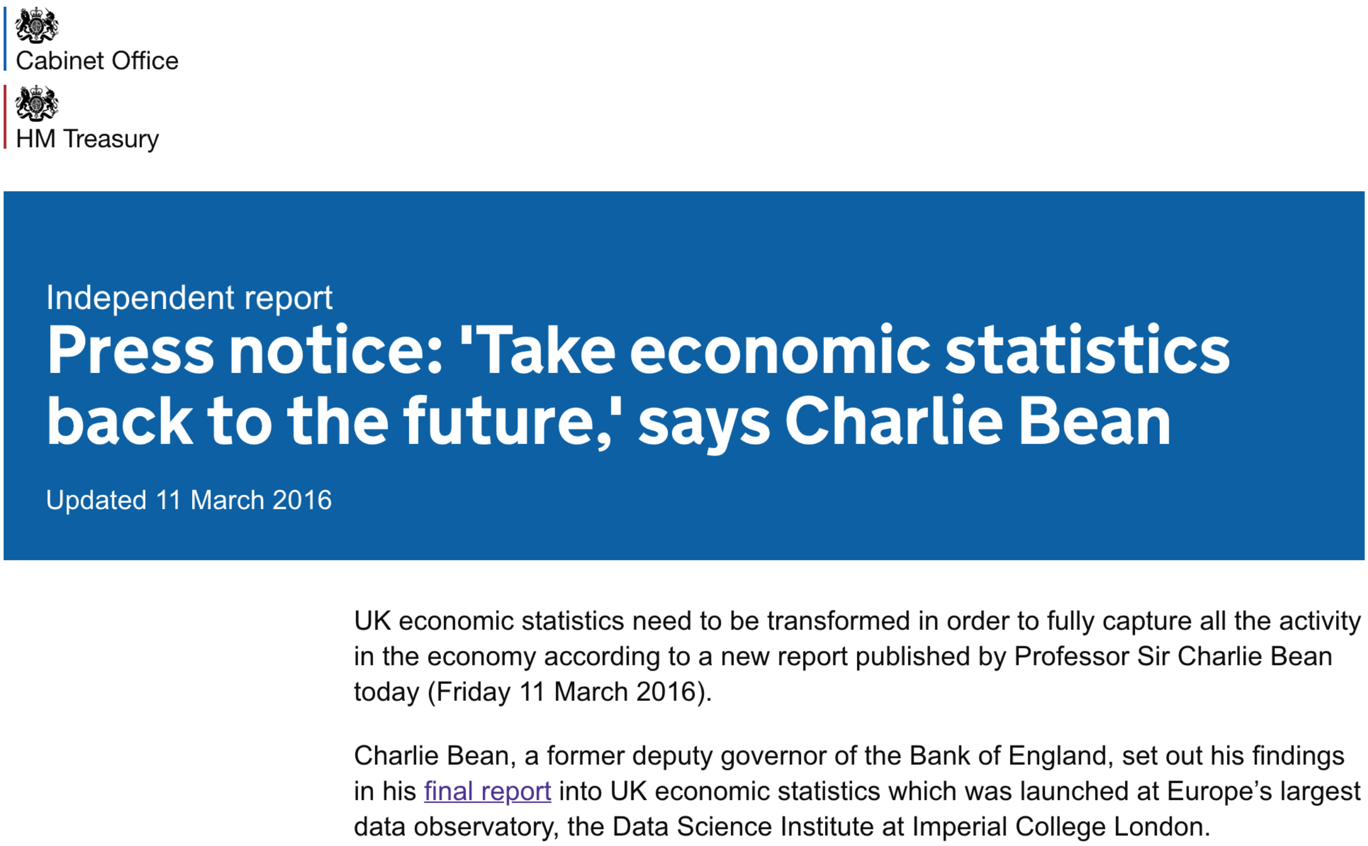


Figure: Conference Organized by Bank of England and US Federal Reserve Board

https://www.bankofengland.co.uk/events/2018/november/modelling-with-big-data-and-machine-learning

Big Data and ML for Policy Makers



Cabinet Office

HM Treasury

Independent report

Press notice: 'Take economic statistics back to the future,' says Charlie Bean

Updated 11 March 2016

UK economic statistics need to be transformed in order to fully capture all the activity in the economy according to a new report published by Professor Sir Charlie Bean today (Friday 11 March 2016).

Charlie Bean, a former deputy governor of the Bank of England, set out his findings in his final report into UK economic statistics which was launched at Europe's largest data observatory, the Data Science Institute at Imperial College London.

https://www.gov.uk/government/publications/independent-review-of-uk-economic-statistics-final-report

Big Data and ML for Policy Makers

ONS Big Data Team

Data Science for Official Statistics

- UK
- Our ONS Page
- Email
- GitHub

Published work

A list of all the Big Data Teams published work. Code repository links where available via Recently added publications are marked with a ⭐

Please contact us at ons.big.data.project@ons.gov.uk if you would like more information about any of the work we have done or are doing!

Methodology working papers

- ⭐ **Špakulová I., Gask K., Hopper, N.A. and James, M.** (2019) Using data science for the address matching service
- ⭐ **Šakulová I., Dove I., Bates, A. and Turner, A.** (2019) Synthetic data pilot

https://onsbigdata.github.io/publications

Big Data and ML in the Industry

J.P.Morgan

May 2017

Big Data and AI Strategies

Machine Learning and Alternative Data Approach to Investing

Google the title to get a copy by yourself.

Table of Contents of Data Providers

In this section we provide a comprehensive list of alternative data and technology providers. In order to navigate this handbook, we provide a table of contents below.

Big Data and ML in Practice

Big Data and ML in Practice

宏观经济大数据平台

主页
数据导入
数据查看
数据治理
数据分析

数据来源	数据类别	条数	字段数量	起始日期
原始爬虫数据	工程建设	9591	14	0001-01-01 00:00:00
原始爬虫数据	政府采购	5876	14	2018-07-30 00:00:00
原始爬虫数据	产权交易	5297	14	2017-10-19 00:00:00
原始爬虫数据	大众点评	81856	17	2019-01-16 14:42:32
原始爬虫数据	前程无忧招聘	9414481	15	2018-12-08 21:34:54
原始爬虫数据	链家新房	677	18	2018-11-20 18:06:08
原始爬虫数据	链家二手房	261980	25	2018-11-27 19:45:45
原始爬虫数据	智联招聘	2302919	15	2018-12-08 16:59:45
互联网数据	百度新闻	571375	6	2018-11-26 03:51:47
互联网数据	佰腾专利	150482	15	2018-11-26 03:56:37
互联网数据	工程建设	9591	11	2018-11-25 04:06:41
互联网数据	大众点评	81856	16	2018-11-25 04:04:48

How Much Does Big Data and ML Really Help in Macroeconomics?

Talk is Cheap, Show Me the Results.

Introduction to Some Empirical Work

Combining Satellite Imagery and Machine Learning to Predict Poverty (2016, Science)

- Background: Nightlight display little variation at lower expenditure levels.
- Research Question: Estimate expenditure and asset wealth with satellite imagery.
- Main Data: survey and satellite data from five African countries.
- Model: Transfer Learning from ImageNet to Daytime Satellite images + ridge regression for prediction.
- Results: explain up to 75% of the variation in local-level economic outcomes.

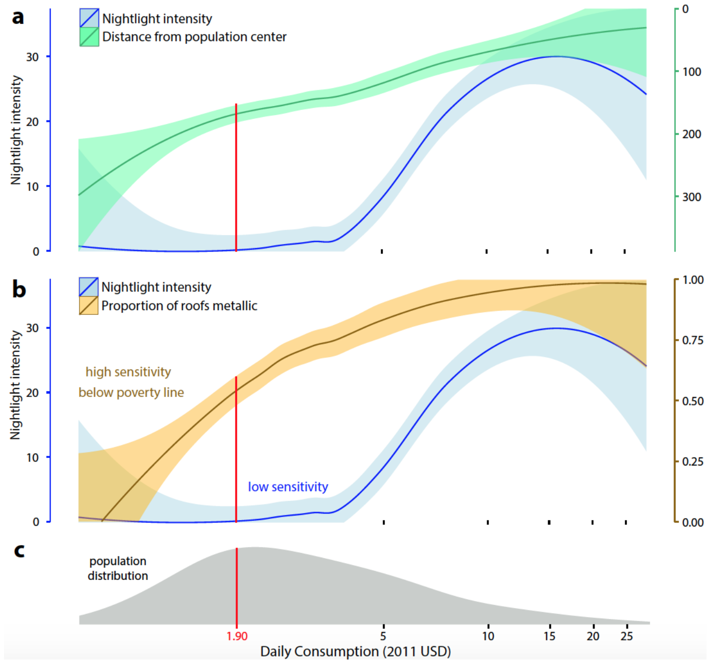


Figure: Consumption vs. Nightlight Intensity and Other Features

Combining Satellite Imagery and Machine Learning to Predict Poverty (2016, Science)

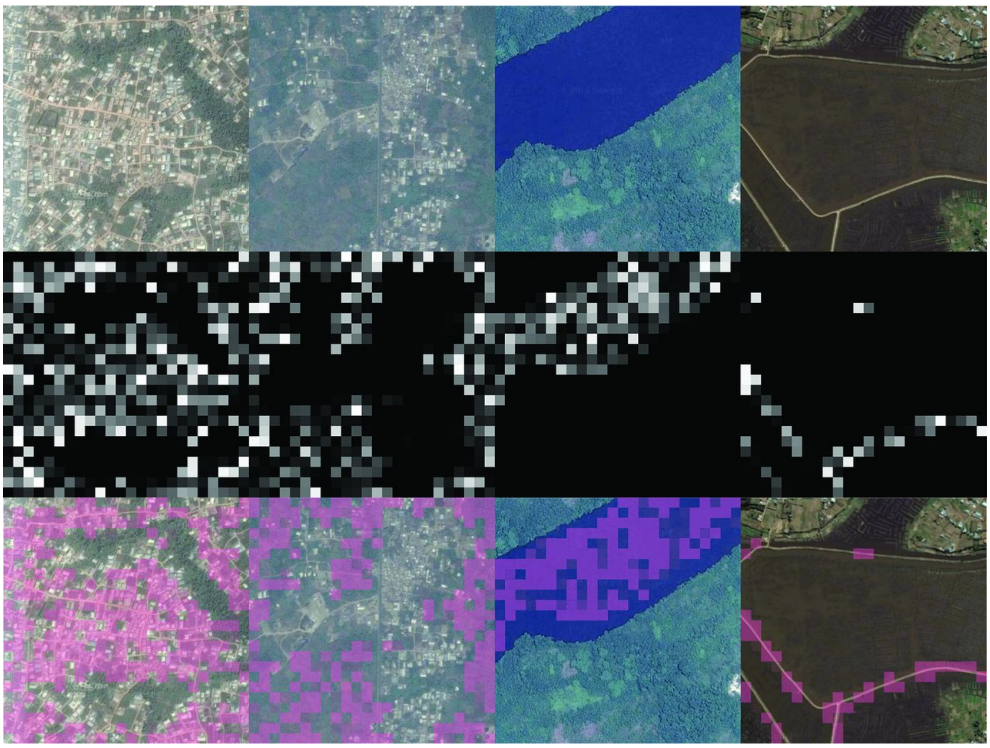


Figure: Feature Extraction from Daytime Satellite Images with Transfer Learning

Combining Satellite Imagery and Machine Learning to Predict Poverty (2016, Science)

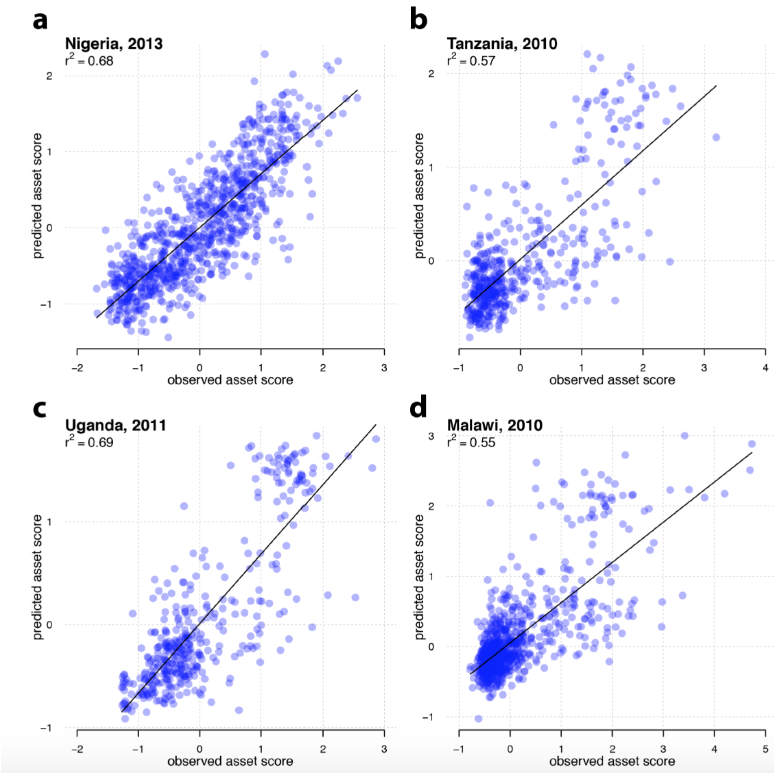


Figure: Cluster-level Asset: TL Based Prediction vs. Survey Measure

Predicting Poverty and Wealth from Mobile Phone Metadata (2015, Science)

- Research Question: Use individual's mobile phone use history to infer his/her socioeconomic status, and reconstruct the distribution and aggregates of wealth of an entire nation.
- Main Data:

Table 1. Summary statistics for primary data sets. Phone survey data were collected by the authors in Kigali, in collaboration with the Kigali Institute of Science and Technology. Call detail records were collected by the primary mobile phone operator in Rwanda at the time of the phone survey. Demographic and Health Survey (DHS) data were collected by the Rwandan National Institute of Statistics. N/A, not applicable.

Summary statistic	Phone survey	Call detail records	DHS (2007)	DHS (2010)
Number of unique individuals	856	1.5 million	7377	12,792
Data collection period	July 2009	May 2008–May 2009	Dec. 2007–Apr. 2008	Sept. 2010–Mar. 2011
Number of questions in survey	75	N/A	1615	3396
Primary geographic units	30 districts	30 districts	30 districts	30 districts
Secondary geographic units	300 cell towers	300 cell towers	247 clusters	492 clusters

- Model: (1) combinatorial method for feature engineering; (2) elastic net for prediction; (3) aggregate.
- Results: see next two pages.

Predicting Poverty and Wealth from Mobile Phone Metadata (2015, Science)

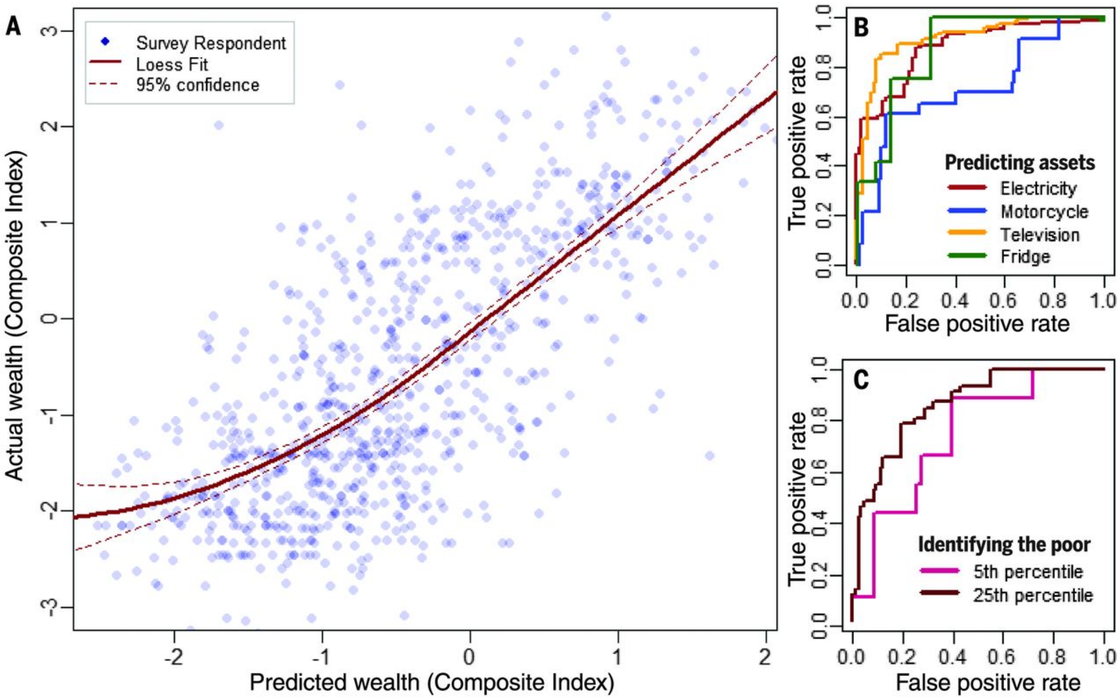


Figure: Predicting survey responses with phone data at individual level

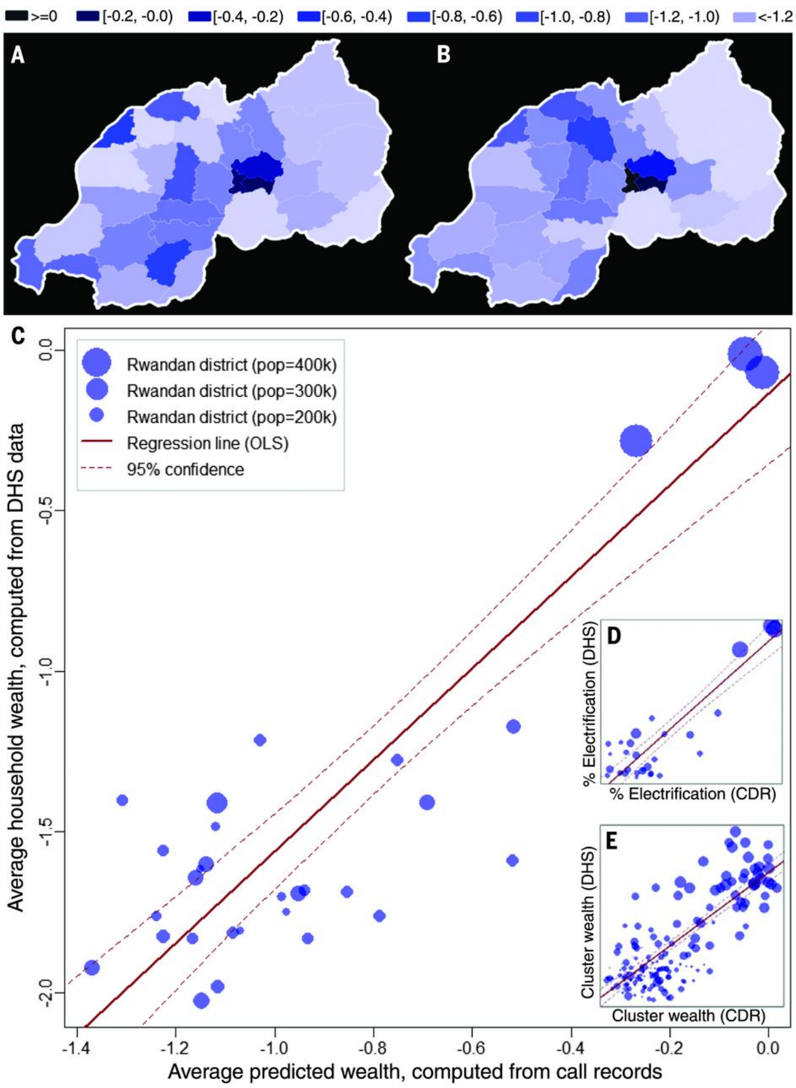


Figure: Comparison of wealth predictions to government survey data at district level

Estimating Unemployment Rate with Administrative Data (E, Huang, Yang, Yang, Zheng, et al., 2019)

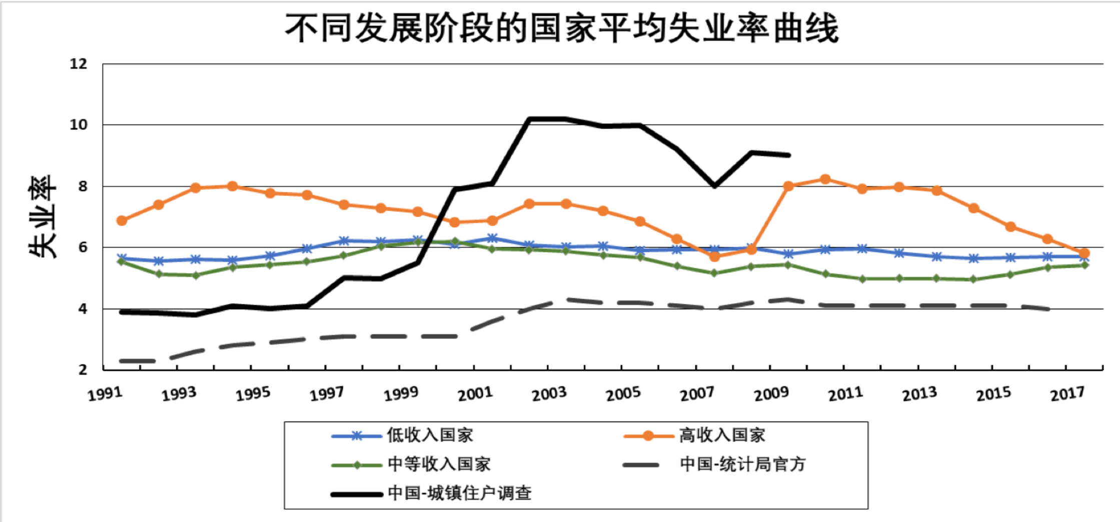


Figure: Replication from Feng et al. (2017)

- Background: China's unemployment data are very uninformative.
- Research Question: Predict individual employment status with administrative data and aggregate up.
- Main Data: administrative data from a city with 4M population.

Estimating Unemployment Rate with Administrative Data

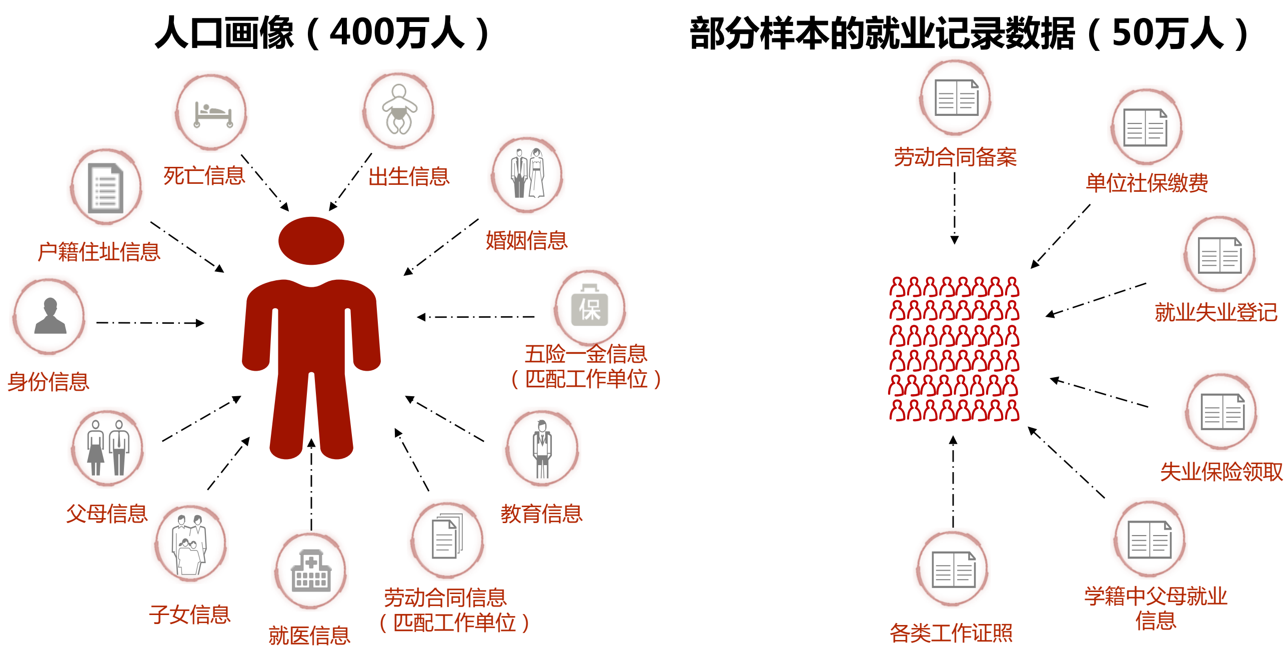


Figure: Left: Construction of Input Variables; Right: Construction of Output Labels. (Source: Beijing Institute of Big Data Research)

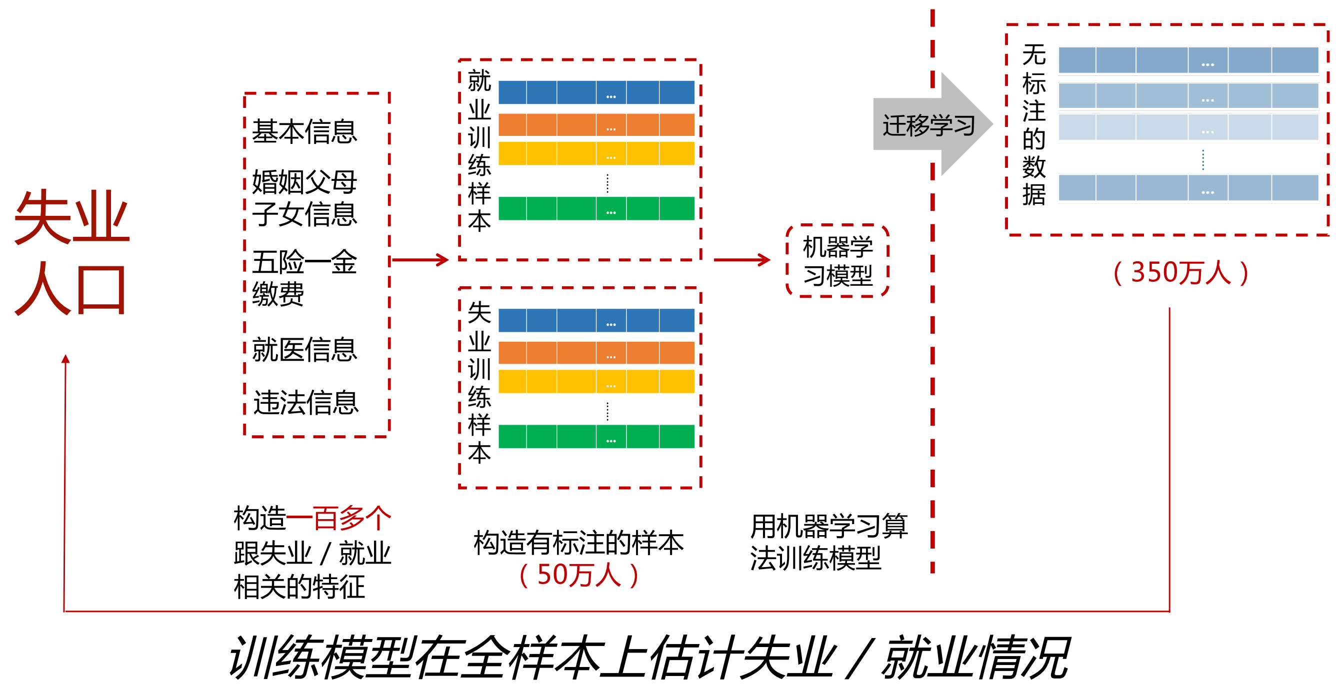
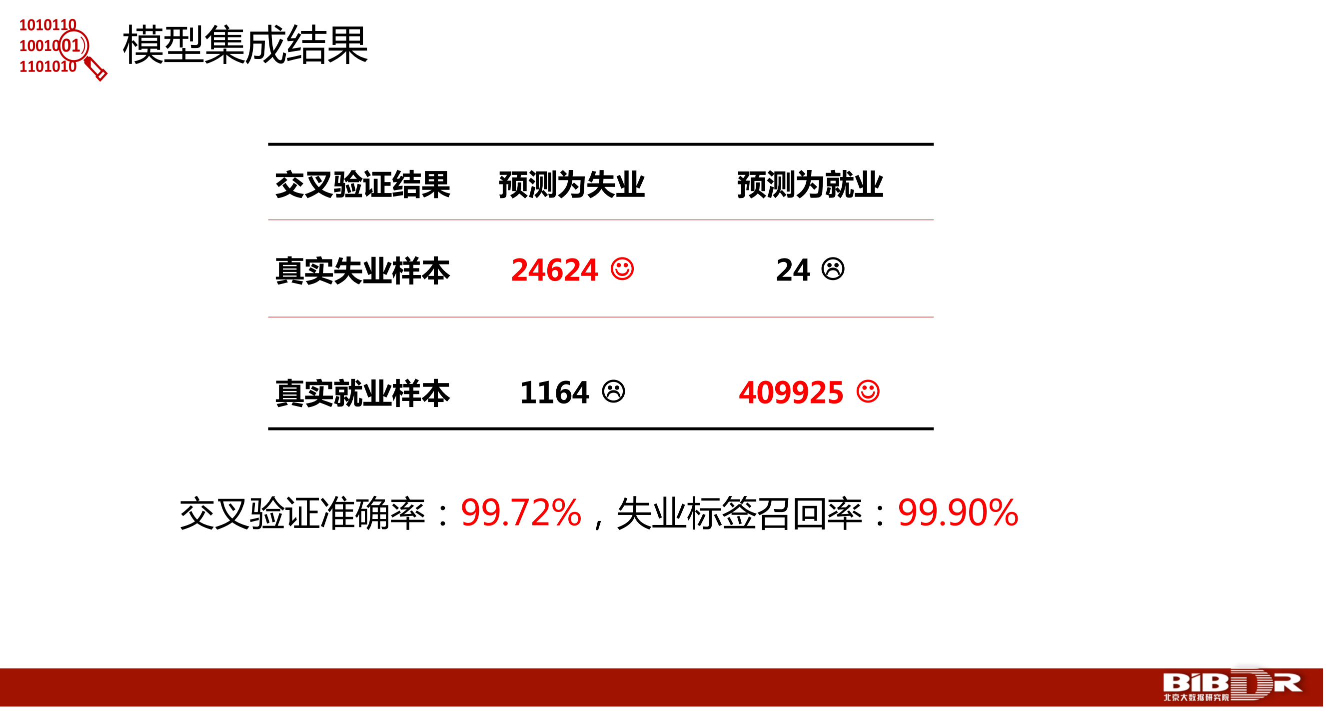


Figure: Model Illustration (Source: Beijing Institute of Big Data Research)

Estimating Unemployment Rate with Administrative Data

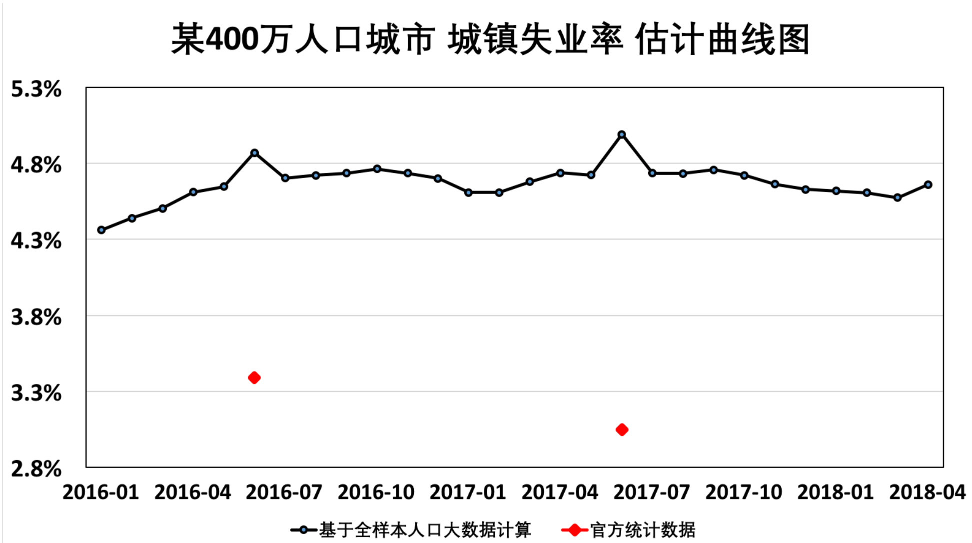

模型集成结果

交叉验证结果	预测为失业	预测为就业
真实失业样本	24624 ☺	24 ☹
真实就业样本	1164 ☹	409925 ☺

交叉验证准确率：99.72%，失业标签召回率：99.90%

Figure: Model Results (Source: Beijing Institute of Big Data Research)

Estimating Unemployment Rate with Administrative Data



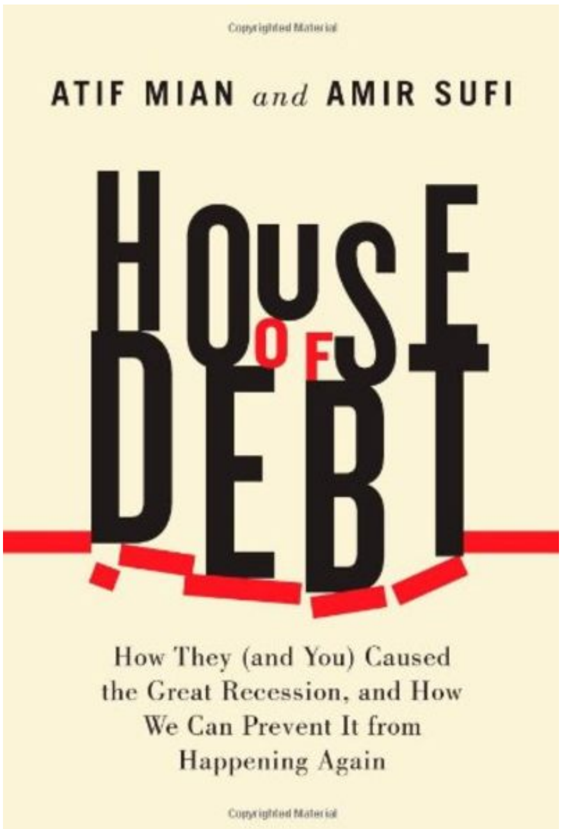
主要发现

- 估计出的失业率曲线有明显周期性特征，数值水平处于合理区间
- 2017年7月的失业高峰与当地经验高度一致

Figure: Model Results (Source: Beijing Institute of Big Data Research)

Mian and Sufi Narrative on the Great Recession

- 2002 to 2006: Housing boom → house prices ↑→ sub-prime mortgage supply ↑ & households leverage ↑.
- 2007 to 2009: Housing bust → house prices ↓→ household wealth ↓.
- In areas with highest leverage and largest ↓ in house prices → consumption ↓ most.
- Due to real frictions, house prices ↓→ housing net worth ↓→ employment in non-tradable industries ↓.



Mian and Sufi Narrative on the Great Recession

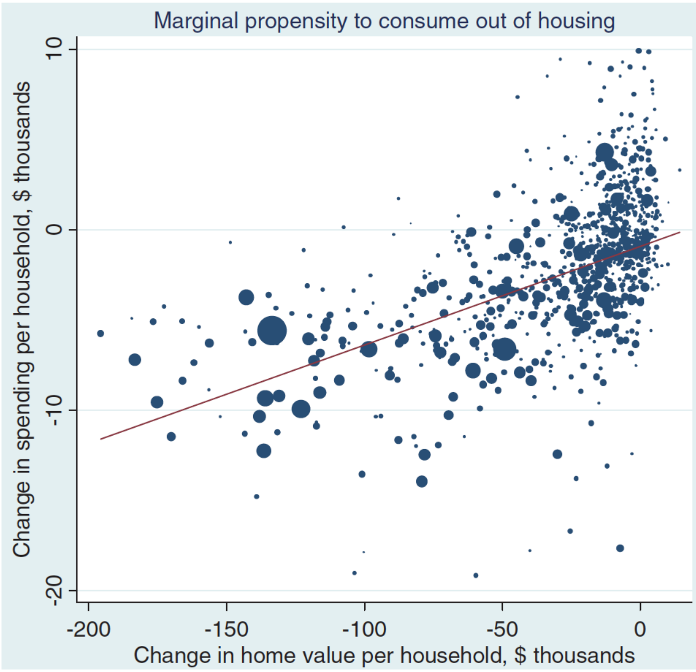


Figure: Change in Home Value vs. Consumption

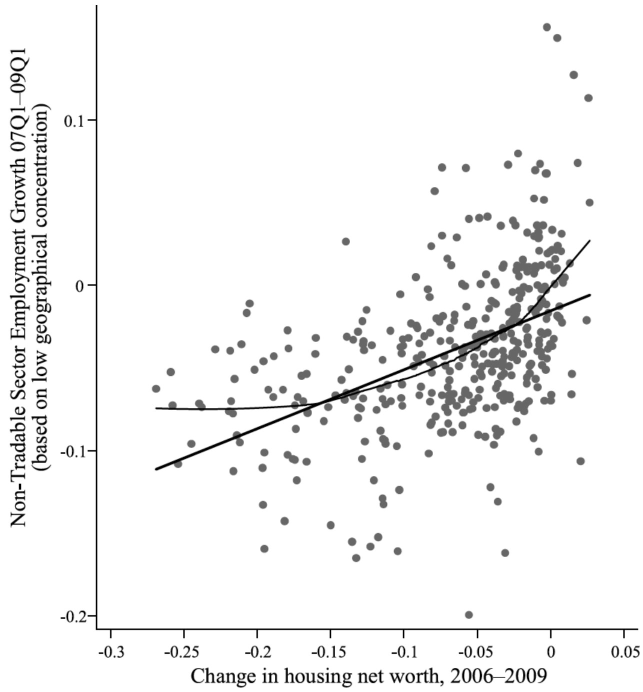


Figure: Change in Home Value vs. Non-tradable sector Employment

Most of the Mian-Sufi narratives are based on empirical work with small regional (eg ZIP code) level data.

Using High-Frequency Detailed Transaction Data to Study Consumption

Baker, S.R., 2018. Debt and the response to household income shocks: Validation and application of linked financial account data. *Journal of Political Economy*, 126(4), pp.1504-1557.

- Research Question: How consumption elasticity vary among households? Through what channels?
- Main Data: A large online personal finance website with 4 million users' (1) Transaction data: time-stamped spending and income records with detailed information (source, category, instrument, etc.) (2) Balance sheet data: daily updated in investment, equity, retirement, real estate, and loan accounts. (3) Demographic data.
- Model: IV regression.
- Results: (1) Elasticity of consumption is significantly higher in households with more debt and fewer assets. (2) Debt is not significant after controlling credit and liquidity constraints.

Using High-Frequency Detailed Transaction Data to Study Consumption

IMPACT OF DEBT AND CREDIT ON ΔLOG(SPENDING) FOLLOWING INCOME SHOCKS

	Sample: All (IV)					
	(1)	(2)	(3)	(4)	(5)	(6)
ΔLog(Income)	.321***	.343***	.346***	.329***	.334***	.324***
	(.032)	(.026)	(.023)	(.022)	(.021)	(.023)
ΔLog(Income) $\times$ [Debt/(Debt+Assets)]	.087***	.073***	.052***	.031**	.024*	.016
	(.026)	(.024)	(.016)	(.015)	(.014)	(.016)
ΔLog(Income) $\times$ (Credit Score)		−.037***	−.030**	−.026**	−.019*	−.026**
		(.014)	(.011)	(.012)	(.011)	(.012)
ΔLog(Income) $\times$ (Unused Credit)			−.062***	−.059***	−.051***	−.043***
			(.012)	(.011)	(.012)	(.011)
ΔLog(Income) $\times$ (Liquid Assets)				−.073***	−.071***	−.068***
				(.015)	(.016)	(.019)
ΔLog(Income) $\times$ (Credit Limit Decline)					.063*	.069*
					(.034)	(.036)
ΔLog(Income) $\times$ (Marginal Interest Rate)						.094**
						(.046)
Observations	3,014,721	3,014,721	3,014,721	3,014,721	3,014,721	3,014,721

Macroeconomic Nowcasting and Forecasting with Big Data

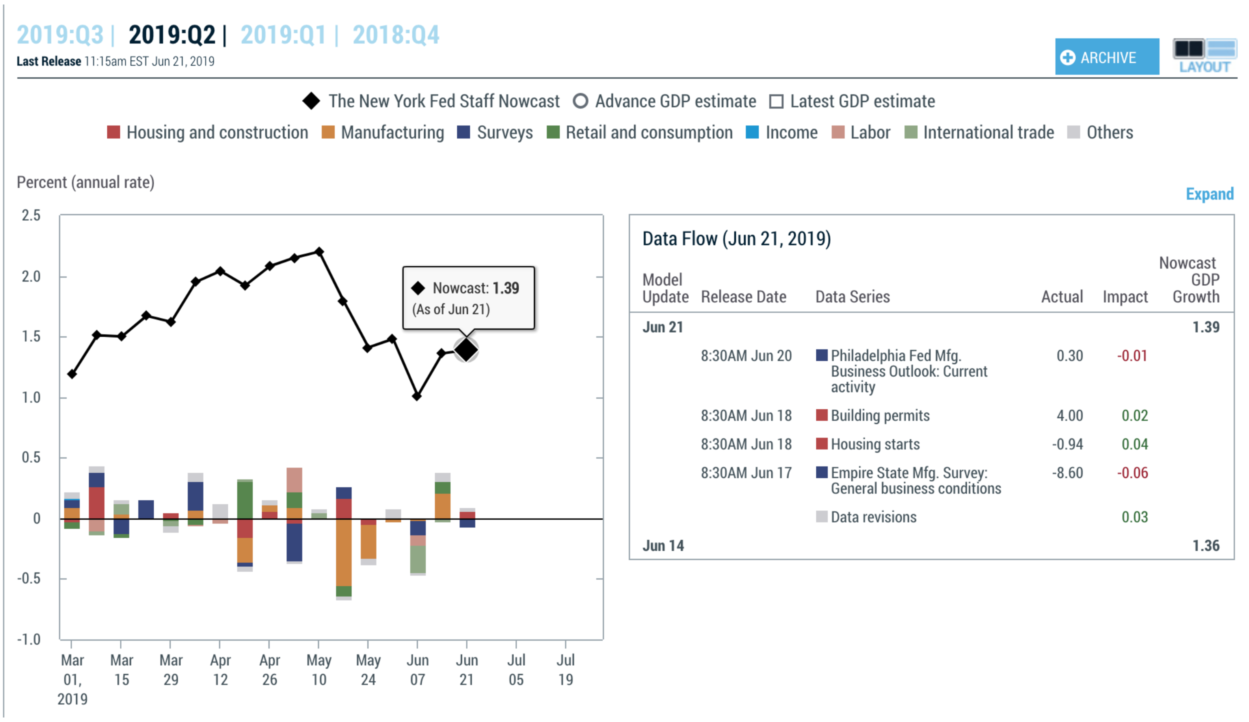


Figure: Nowcasting GDP growth with a wide range of macroeconomic data.

https://www.newyorkfed.org/research/policy/nowcast

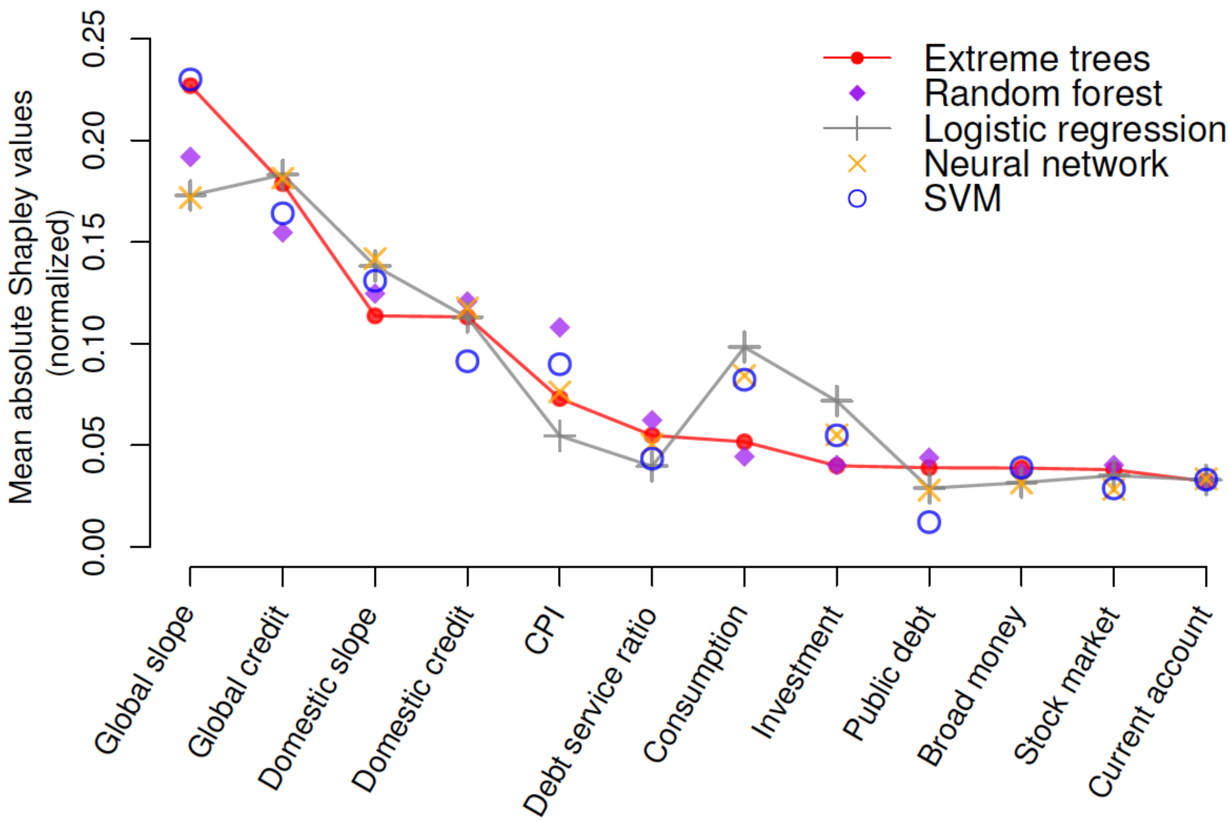
Bok, B., D. Caratelli, D. Giannone, A. Sbordone, and A. Tambalotti. 2017. Macroeconomic Nowcasting and Forecasting with Big Data. Federal Reserve Bank of New York *Staff Reports*, no. 830, November.

Predicting Financial Crisis with Machine Learning

Bluwstein et al. (2019): by Economists at the Bank of England.

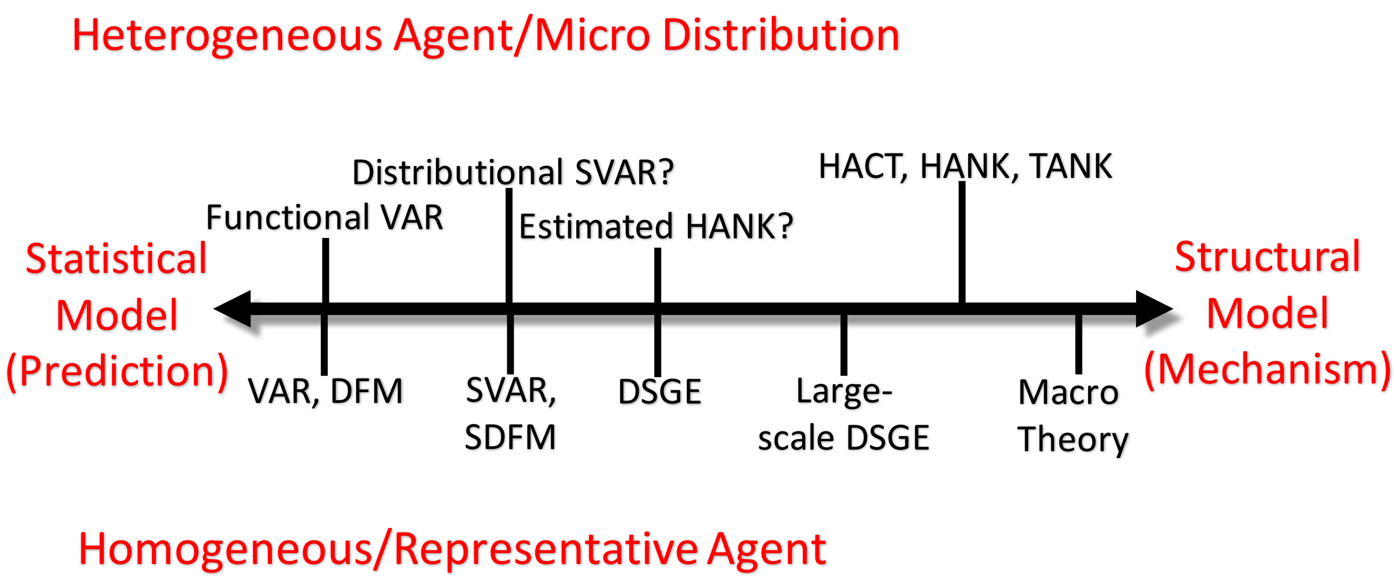
- Research Question: Build early warning models for financial crises using ML techniques and identify the key drivers of financial crises.
- Main Data: macroeconomic and financial time series data set for 17 countries between 1870-2016 (Jordà-Schularick-Taylor Macrohistory Database).
- Model: (1) ML models for prediction; (2) Shapley value framework for key driver detection.
- Results: (1) ML models outperform linear or logistic regressions; (2) The most important predictors are the slope of the yield curve and credit growth, and the importance order is persistent in different models.

Variable Importance for Crisis Prediction in Different ML Models

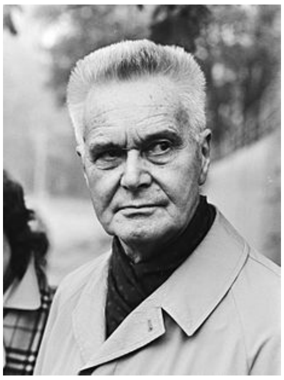


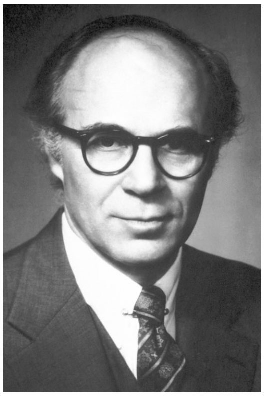
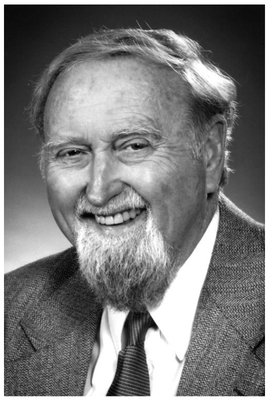
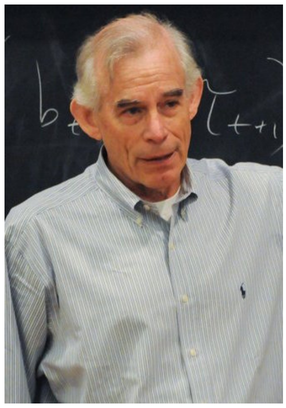
Introduction to Some Methodological Work

The Spectrum of Macro Methodology



Statistical Approach of Macroeconomics



- Pioneered by Nobel laureates Jan Tinbergen (1939), Trygve Haavelmo (1943), Lawrence Klein et al. (1965, 1969), Clive Granger (1969), Christopher Sims (1972, 1980).
- Vector Autoregressive (VAR) Model: $Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + \epsilon_t$. Natural extension to machine learning models for time series.
- Structural VAR:

$$B_0 Y_t = \sum_{i=1}^{p} B_i Y_{t-i} + \eta_t$$

so that different dimensions of η_t are independent.

Dynamic Factor Model (DFM): Major Big Data Tool for Macroeconomists

- DFM with $N \times 1$ observable X_t to $p \times 1$ factor F_t ($N \gg p$):

$$X_t = C + \Lambda F_t + e_t, e_t \overset{iid}{\sim} \mathcal{N}(0, R)$$
$$F_t = \Phi(L) F_{t-1} + u_t, u_t \overset{iid}{\sim} \mathcal{N}(0, Q)$$

- Particularly useful when input data X_t is high dimensional. Can handle mixed frequency input.
- Low dimensional factors F_t can be used for measurement, nowcasting, forecasting, etc.
- Special case of state space models, or hidden Markov models, which is closely connected to ML models like RNN and LSTM.

Structural Approach of Macroeconomics

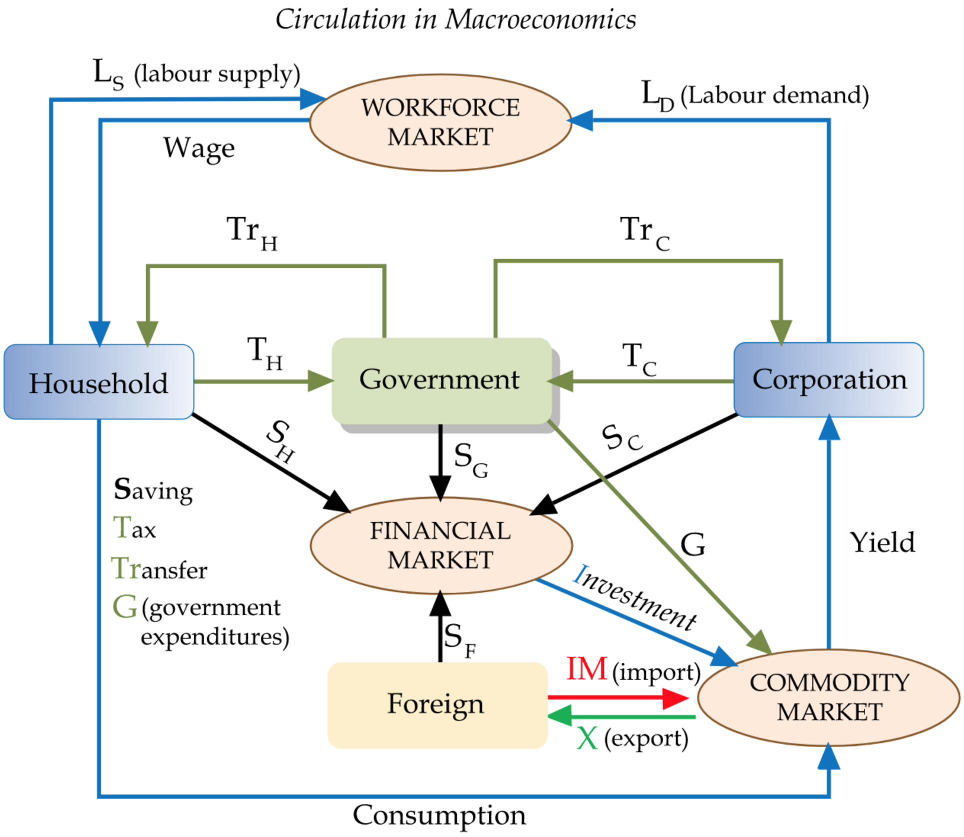


Figure: Illustration of Structural Models. (source: Wikipedia)

Krusell and Smith (1998)

- Uninsurable idiosyncratic income risk → heterogeneous agents.
- Distribution of agents' wealth Γ_t come into the state space → computationally hard.

$$v(k_t, \varepsilon_t; \Gamma_t, z_t) = \max_{c_t, k_{t+1} \geq 0} u(c_t) + \beta E_t [v(k_{t+1}, \varepsilon_{t+1}; \Gamma_{t+1}, z_{t+1})]$$

s.t.

$$c_t + k_{t+1} = r(K_t, L_t, z_t) k_t + w(K_t, L_t, z_t) \bar{e} \varepsilon_t + (1 - \delta) k_t$$

$$\Gamma_{t+1} = H(\Gamma_t, z_t, z_{t+1})$$

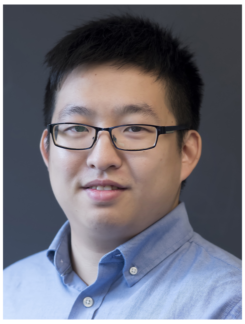
- Krusell-Smith method: "approximate aggregation". Capture the whole distribution with limited moments $\Gamma_t \approx \{m_1, m_2, ..., m_M\}_t$.
- Conclusion: heterogeneity does not matter, and the first moment accounts for most variation.

HANK: Heterogeneous Agent New Keynesian Model

Monetary policy according to HANK
G Kaplan, B Moll, GL Violante - American Economic Review, 2018 - aeaweb.org
We revisit the transmission mechanism from monetary policy to household consumption in a Heterogeneous Agent New Keynesian (HANK) model. The model yields empirically realistic distributions of wealth and marginal propensities to consume because of two features …

- One of the most popular methodological advances in macro recently.
- Building blocks of HANK:
    1. Uninsurable idiosyncratic income risk (HA).
    2. Nominal price rigidities (NK).
    3. Assets with different degrees of liquidity (HtM).
    4. Continuous time approach (HACT).
- Contribution: New framework for quantitative analysis of the transmission mechanism of monetary policy that matches data much better.

HACT: The Hero behind HANK


Yves Achdou, Jiequn Han, Jean-Michel Lasry, Pierre-Louis Lions and Benjamin Moll. Income and Wealth Distribution in Macroeconomics: A Continuous-Time Approach. revise & resubmit, *Review of Economic Studies*.

- Formulate heterogeneous agent models as coupled PDEs - mean field games (MFG).
- Efficient computational algorithms are available.
- New theoretical results.

HACT: PDE Formulation

- Hamilton-Jacobi-Bellman equation:

$$\rho v_i(a,t) = \max_c u(c) + \partial_a v_i(a,t)\left[z_i + r(t)a - c\right] \\ +\lambda_i\left[v_j(a,t) - v_i(a,t)\right] + \partial_t v_i(a,t)$$

  with terminal condition $v_i(a,T) = v_{i,\infty}(a)$.
- Kolmogorov Forward equation:

$$\partial_t g_i(a,t) = -\partial_a\left[s_i(a,t)g_i(a,t)\right] - \lambda_i g_i(a,t) + \lambda_j g_j(a,t)$$

  for $i = 1,2$ and $j \neq i$ with initial condition $g_i(a,0) = g_{i,0}(a)$.
- State constraint boundary condition:

$$u'\left(z_i + r(t)\underline{a}\right) \geq \partial_a v_i(\underline{a},t), \quad i = 1,2.$$

Deep Learning: Possible Tool to Solve HA Models

DL to solve stochastic dynamic programming:

- Han, Jiequn and E, Weinan, 2016. Deep learning approximation for stochastic control problems. *NIPS Workshop*.

DL to solve high-dimensional PDEs:

- Han, Jiequn, Jentzen, A. and E, Weinan, 2018. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34), pp.8505-8510.

DL to solve the Krusell-Smith problem:

- Fernández-Villaverde, J., Hurtado, S. and Nuno, G., 2019. Financial Frictions and the Wealth Distribution.

DL to solve continous time DSGE models:

- Duarte, V., 2018. Machine Learning for Continuous-Time Finance. revise & resubmit, *Review of Financial Studies*.

Deep Learning for Stochastic Control

- Look for a feedback control: $a_t = a_t(s_t)$ to

$$\min_{\{a_t\}_{t=0}^{T-1}} \mathbb{E}\Big\{ \sum_{t=0}^{T-1} c_t(s_t, a_t(s_t)) + c_T(s_T) \mid s_0 \Big\}$$

- Traditional methods in operation research: discretize state and/or control into finite spaces + approximate dynamic programming.
- Neural network approximation:

$$a_t(s_t) \approx a_t(s_t|\theta_t),$$

Solve directly the approximate optimization problem

$$\min_{\{\theta_t\}_{t=0}^{T-1}} \mathbb{E}\Big\{ \sum_{t=0}^{T-1} c_t(s_t, a_t(s_t|\theta_t)) + c_T(s_T) \Big\},$$

rather than dynamic programming principle.

Deep Learning for Stochastic Control

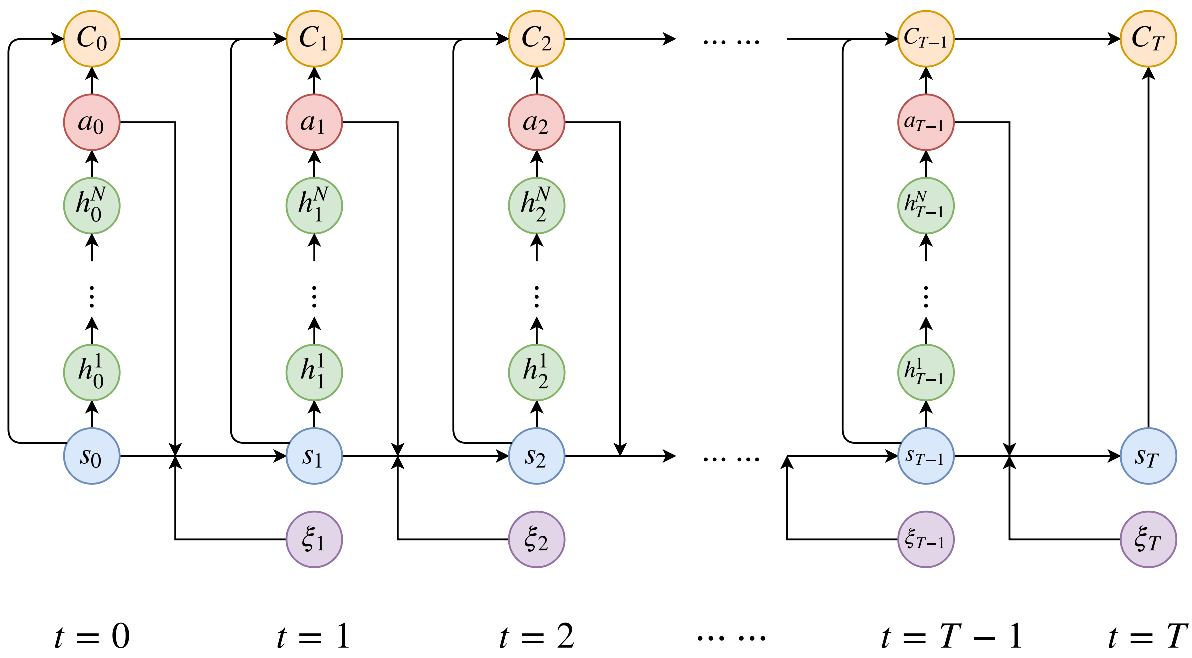


Figure: Network architecture for solving stochastic control in discrete time (Han and E, 2016). The whole network has $(N+1)T$ layers in total that involve free parameters to be optimized simultaneously. Each column (except ξ_t) corresponds to a sub-network at t.

Deep Learning for Semilinear Parabolic PDEs

Consider a general semilinear parabolic PDE in $[0, T] \times \mathbb{R}^d$:

$$\frac{\partial u}{\partial t}(t,x) + \frac{1}{2}\left(\sigma\sigma^{\mathsf{T}}(t,x)(\mathrm{Hess}_x u)(t,x)\right) + \nabla u(t,x)\cdot\mu(t,x) \\ + f\left(t,x,u(t,x),\sigma^{\mathsf{T}}(t,x)\nabla u(t,x)\right) = 0.$$

The terminal condition $u(T,x) = g(x)$ is given. Under suitable regularities, given a stochastic process satisfying

$$X_t = \xi + \int_0^t \mu(s,X_s)\,ds + \int_0^t \sigma(s,X_s)\,dW_s,$$

the solution of PDE satisfies the following SDE

$$\begin{aligned} &u(t,X_t) - u(0,X_0) \\ &= -\int_0^t f\left(s,X_s,u(s,X_s),\sigma^{\mathsf{T}}(s,X_s)\,\nabla u(s,X_s)\right) ds \\ &\quad + \int_0^t [\nabla u(s,X_s)]^{\mathsf{T}}\,\sigma(s,X_s)\,dW_s. \end{aligned}$$

Deep Learning for Semilinear Parabolic PDEs

Time Discretization:

$$X_{t_{n+1}} - X_{t_n} \approx \mu(t_n, X_{t_n})\,\Delta t_n + \sigma(t_n, X_{t_n})\,\Delta W_n,$$

$$\begin{aligned} &u(t_{n+1}, X_{t_{n+1}}) - u(t_n, X_{t_n}) \\ &\approx -f\left(t_n, X_{t_n}, u(t_n, X_{t_n}), \sigma^{\mathsf{T}}(t_n, X_{t_n})\,\nabla u(t_n, X_{t_n})\right)\Delta t_n \\ &\quad + [\nabla u(t_n, X_{t_n})]^{\mathsf{T}}\,\sigma(t_n, X_{t_n})\,\Delta W_n, \end{aligned}$$

Key step: approximate the function $x \mapsto \sigma^{\mathsf{T}}(t,x)\,\nabla u(t,x)$ at each discretized time step $t = t_n$ by a feedforward neural network

$$\begin{aligned} \sigma^{\mathsf{T}}(t_n, X_{t_n})\,\nabla u(t_n, X_{t_n}) &= (\sigma^{\mathsf{T}}\nabla u)(t_n, X_{t_n}) \\ &\approx (\sigma^{\mathsf{T}}\nabla u)(t_n, X_{t_n}|\theta_n), \end{aligned}$$

where θ_n denotes neural network parameters.

Deep Learning for Parabolic PDEs: Similar Structure

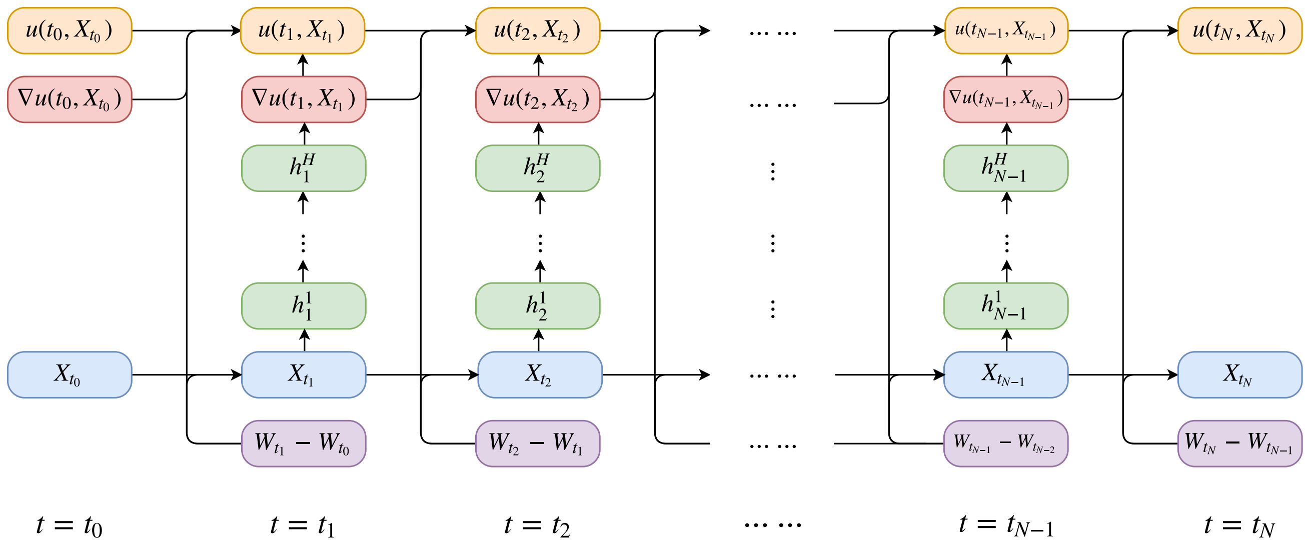


Figure: Network architecture for solving parabolic PDEs (Han, Jentzen and E, 2016). Each column corresponds to a subnetwork at time $t = t_n$. The whole network has $(H + 1)(N - 1)$ layers in total that involve free parameters to be optimized simultaneously.

SUMMARY: WHAT CAN BIG DATA AND ML BRING TO MACRO?

Summary: What can Big Data and ML Bring to Macro?

- New Data Sources:
  1. More important role for micro big data in macro research, especially granular level administrative data and proprietary data on households, firms, small areas, etc..
  2. Alternative data as proxies, instruments or data input.
  3. Use ML to generate new data (eg. textual data, image data).
- New Modeling Tools:
  1. Tools to handle big data. New tools for both microeconometrics and macroeconometrics.
  2. Tools to handle micro-founded high dimensional models.
- New Knowledge System:
  1. New indicator system.
  2. New knowledge of macro.

Organization of the Course

Administrative Information

- Quite intense course: we meet (at least) three times every week.
- A course with lots of new stuff - we also learn by teaching.
- Lectures: T, Th 9:00 - 12:00.
- In-class presentations: F 10:00 - 12:00 from the second week.
- Grading: 50% class participation and presentation. 50% final project.
- Auditing students are welcome, but should also participate in the in-class presentation.
- All the feedbacks are welcome!

Organization of Lectures

- Overview
  1. Introduction
  2. Basics of Machine Learning for Macroeconomics
- Statistical Model in Macroeconomics and Machine Learning
  1. Vector Autoregressive Model and Structural VAR
  2. State Space Model, Filtering Problem and EM Algorithm
  3. Recurrent Neural Network and LSTM Network
  4. State Space Model with Non-standard Data
- Structural Model in Macroeconomics and Machine Learning
  1. Representative Agent Model and DSGE
  2. Heterogeneous Agent Model: Krusell-Smith and Variants
  3. Heterogeneous Agent Model in Continuous Time: HACT and HANK
  4. Solving High-dimensional Stochastic Control and PDEs using Deep Neural Networks
  5. Solving Structural Model using Deep Neural Networks

In-Class Presentation

- 8 potential topics are posted on the syllabus.
- Almost all the papers listed use big data to address important macroeconomic questions, published or to be published in decent journals.
- Each group pick up one core paper to present, and are encouraged to talk about other relevant papers under that topic.
- Each group send me the group member names and at least top 3 choices of papers by 6 pm this Thursday.
- Each group present for 30 minutes (including Q&A).
- Presenters should meet the instructor at least 3 days in advance (on or before Tuesday) to talk about the key points of the paper.
- Requirements: clear presentation on the paper's motivation, research question, institutional background, data, empirical specification, empirical findings, model setup, model results, implications, conclusion, with your own critical discussion.

Final Project

- Two options for the final project: a detailed 5-page proposal with preliminary results on a relevant original research; or a replication of one of the method papers covered in the class.
- Potential projects for original research would be discussed by the instructors in class.
- Students should meet the instructor to talk about the choice of the final project before the third week of class (July 14).

Macroeconomic Analysis with Machine Learning and Big Data

Lecture 2: Supervised Learning

Instructor: Weinan E *Scribe: Guanhua Huang, Zhong Zheng*

1 Introduction to Machine Learning

- **Statistics**: less focus on algorithm, more model design & hypothesis testing)
- **Machine Learning**: less focus on hypothesis testing, more on algorithm
  - **supervised learning**:
    - **Data**: $\{x_j, y_j\}_{j=1}^n$
    - Model: $y_j = f^*(x_j) + \epsilon_j$ where y_j is label, ϵ_j is noise.
    - **Objective**: $approximate f^*$
  - **unsupervised learning**:
    - **Data**: $\{x_j\}_{j=1}^n$
    - **Model**: $y_j = f^*(x_j) + \epsilon_j$ where y_i is label
    - **Objective**: find out the rule
  - **reinforcement learning**: policy function: from state space to action space
    - **Objective**: optimal decision

2 Supervised Learning

- **Regression**: $f^* : D \subset R^d \to R$ where $f*$ is continuous.
- **Classification**: $f^* : D \to G(finite\ set), G = \{-1, 1\}$
- **Framework**
  - **Hypothesis Space** $\mathcal{H}_m$
    - e.g. $\mathcal{H}_m = \{w_0 + w^\top x, w_0 \in R, w \in R^d\}$
    - e.g. $\mathcal{H}_m = \{\sum_{j=1}^m \alpha_j \phi_j(x)\}$ where $\{\phi_j(x)\}_{j=1}^m$ is fixed set of function. For example, we can set $\phi_j(x) = cos(k_j x)$
    - e.g. $\mathcal{H}_m = \{\sum_{j=1}^m a_j \sigma(b_j^\top x + c_j)\}$ For example, we can set $\sigma(x) = \max\{0, x\}$(ReLU, an activation function for Neural Network)
    - In examples, m represents (scale of) degree of freedom(You can search **VC dimension** if you want know more about it)

- **Objective Function**: (Loss Function): loss function here is an example of square loss

$$\hat{R}_n(\theta) = \frac{1}{n}\sum_{j=1}^{n}(\hat{y}_j - f(x_j, \theta))^2 + \lambda||\theta||$$

  Where θ is parameter, $||\theta||$ is norm, $\lambda||\theta||$ is regularization term.
- **Optimization Method**
  - Gradient Decent
  - SGD, Adam
  - BFGS .etc

3 Examples for Supervised Learning

3.1 Linear Model

$$\mathcal{H}_m = \{w_0 + w^\top x, w_0 \in R, w \in R^d\}$$

$$\text{write } w_0 + w^\top x \text{ as } w^\top x$$

$$\hat{R}_n(\theta) = \frac{1}{n}\sum_{j=1}^{n}(w^\top x_j - y_j)^2$$

$$\nabla_\theta \hat{R}_n(\theta) = \sum (w^\top x_j - y_j)x_j = 0$$

$$X = (x_1, \ldots, x_n)$$

$$\hat{w} = (XX^\top)^{-1}Xy$$

Regularization **Ridge Regression**

$$\hat{R}_n(\theta) = \frac{1}{n}\sum_{j=1}^{n}(w^\top x_j - y_j)^2 + \lambda||w||^2$$

$$\hat{w} = (XX^\top + \lambda I)^{-1}Xy$$

$$\lim_{\lambda\to 0}(XX^\top + \lambda I)^{-1} = (XX^\top)^{-1}(Generalized\ inverse\ matrix)$$

Use another regularization term

$$\hat{R}_n(\theta) = \frac{1}{n}\sum_{j=1}^{n}(w^\top x_j - y_j)^2 + \lambda N(w)$$

$$N(w) = number\ of\ non-zero\ component\ of\ w$$

Using $\|w\|_1$, Model become **Lasso Regression**

$$\hat{R}_n(\theta) = \frac{1}{n}\sum_{j=1}^{n}(w^\top x_j - y_j)^2 + \lambda||w||_1$$

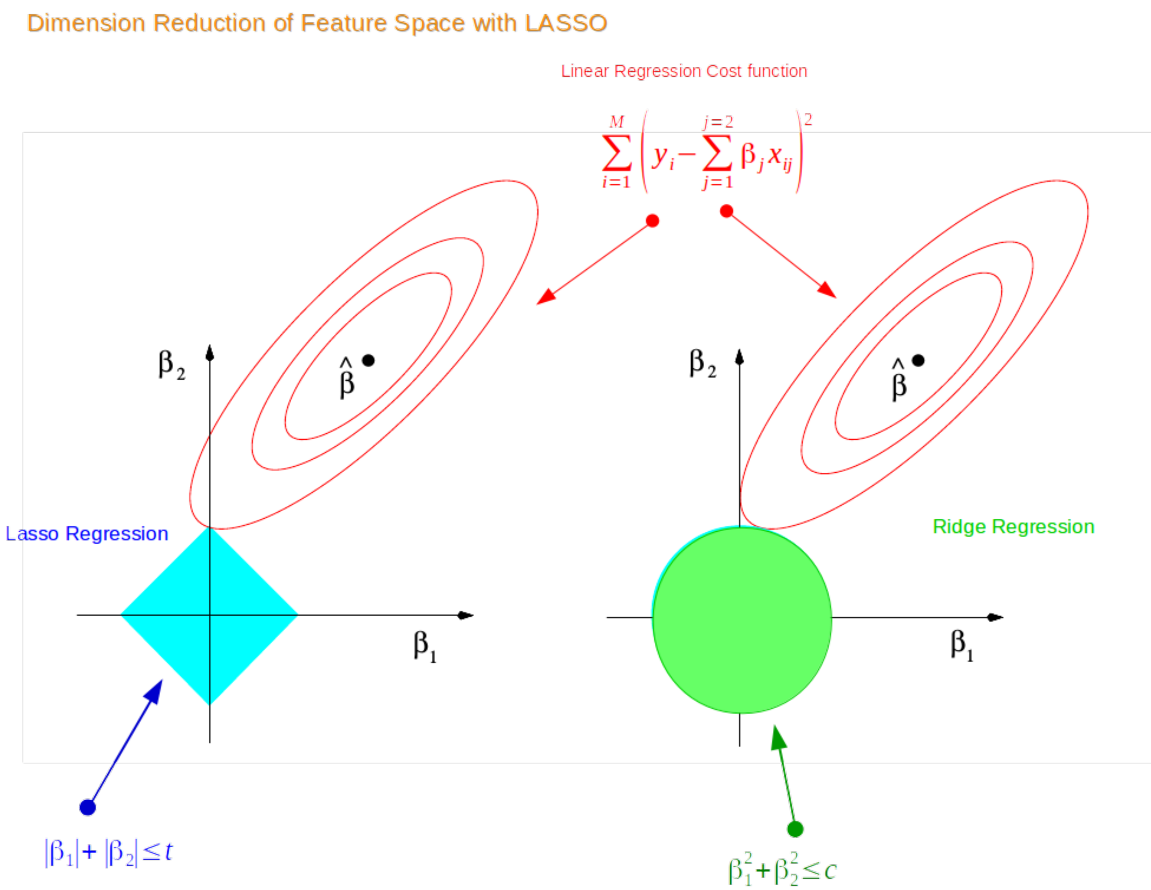


Figure source: `https://towardsdatascience.com`

3.2 Kernel Method

- **kernel function** $k(x, y), x, y \in R^d$, e.g.

$$k(x, y), x, y \in R^d$$

- **Definition**:
  - k is symmetric $k(x, y) = k(y, x)$
  - $\forall \{x_j\}_{j=1}^n \; K = (k(x_i, x_j))_{n\times n} \geq 0$ K is **SPD**(symmetric positive definite)
- **Kernel Space**:

$$\mathcal{H}_m = \{\sum_{j=1}^{n} \alpha_j k(x_j, x)\} \; (m = n)$$

- **Feature-based method**: $\{\phi_j(x)\}_{j=1}^m$ is a set of features

$$\mathcal{H}_m = \{\sum_{i=1}^m \alpha_j \phi_j(x)\}$$

3.3 Neural Network

$$\mathcal{H}_m = \{\sum_{j=1}^m a_j \sigma(b_j^\top x)\}$$

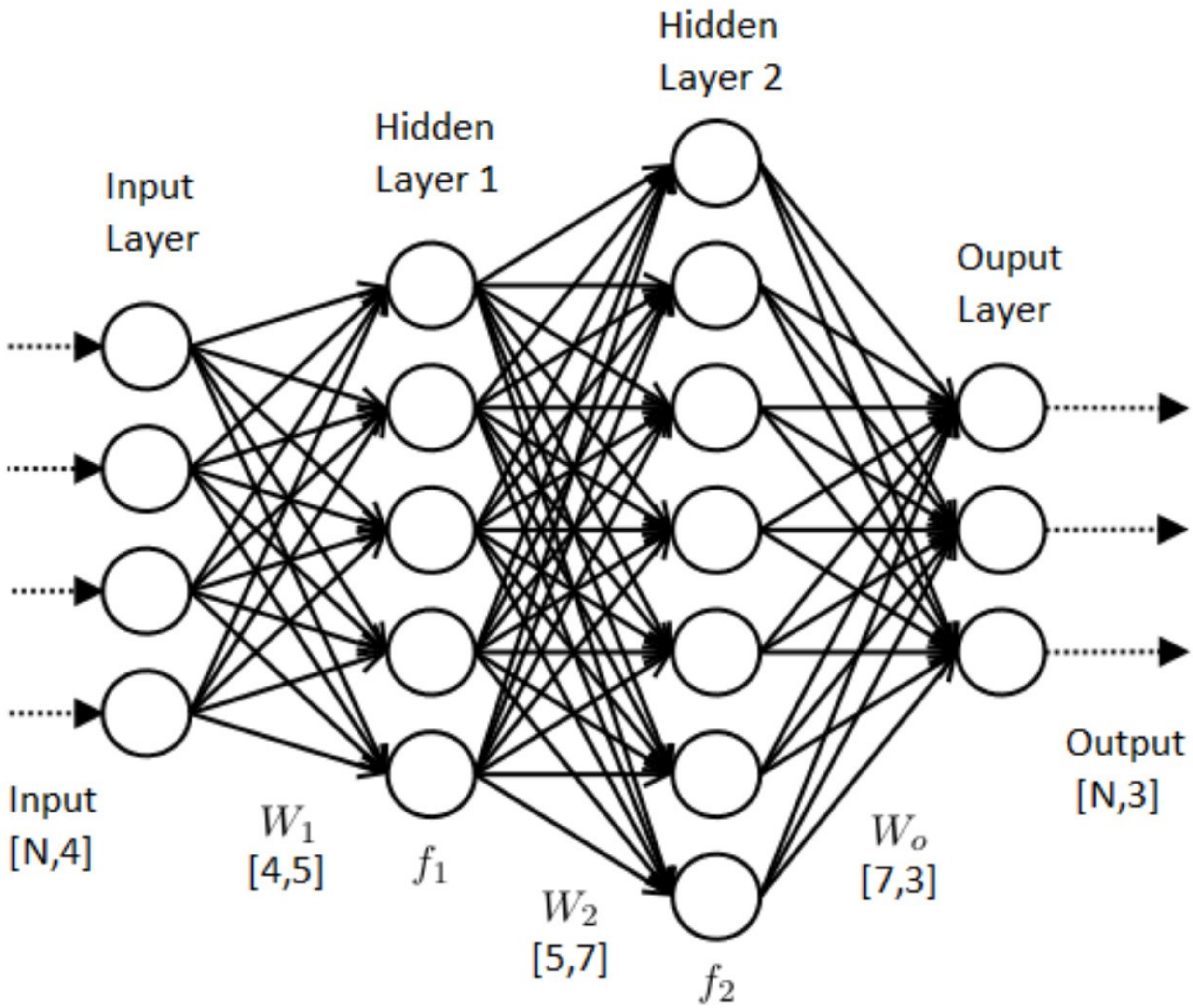


Figure source: `https://www.datasciencecentral.com/profiles/blogs`

Why NN better? (No complete theory in mathematics)

Compared with generalized linear model(GLM)

$$\mathcal{H}_m = \{\sum_{i=1}^{m} \alpha_j \phi_j(x)\}$$

We have a "**Theorem**", For GLM

$$f : R^d \to R,\ d >> 1 \inf_{f_m \in \mathcal{H}_m} \|f_m - f\| \geq cm^{-\frac{1}{d}}$$

For NN

$$f : R^d \to R,\ d >> 1 \inf_{f_m \in \mathcal{H}_m} \|f_m - f\| \leq cm^{-\frac{1}{2}}$$

So let $error = 0.1$. For NN

$$m \sim 10^2$$

For GLM

$$m \sim 10^d$$

3.4 Optimization Algorithm

Gradient Descent:

$$\min_{\theta} F(\theta) \theta_{k+1} = \theta_k - \eta_k \nabla F(\theta_k)$$

For example

$$\nabla F(\theta) = \frac{1}{2} \sum_{j=1}^{n} \nabla (f(\theta, x_j) - y_j)^2 = \sum (f(\theta, x_j) - y_j) \nabla_\theta f$$

Back Propagation : Just use **Chain Rule** In practice, we use Stochastic Gradient Descent (SGD) Methods because of the large scale of data.

3.5 Classification

$$y = \{-1, 1\}$$

$$y = H(f(x)) H(z) = \begin{cases} 1, z > 0 \\ 0, z \leq 0 \end{cases}$$

For continuity, we can let $H(z) = \frac{1}{1+e^{-z}}$(sigmoid)
Logistic Regression: set $f = w^\top x$ and $y = \frac{1}{1+e^{-w^\top x}}$ Above are binary classifications. For **multi-class classification** (K classes), we use softmax

$$q_j(x) = \frac{e^{f_j(x)}}{\sum_{k=1}^{K} e^{f_k(x)}} \sum_{j=1}^{K} q_j(x) = 1$$

Where $q_j(x)$ can be regard as the probability of $x \in class_j$

Macroeconomic Analysis with Machine Learning and Big Data

Lecture 3: Unsupervised Learning

Instructor: Weinan E *Scribe: Guanhua Huang, Lu Yang*

1 Introduction

- **Data**: $\{x_j\}_{j=1}^n$ no label
- **Tasks of unsupervised learning:**
  - Clustering
  - Dimension Reduction
  - Density Estimation

2 Clustering

- **Model**: $x_j = \bigcup_{k=1}^K c_k, c_p \bigcap c_q =, \forall p \neq q$ where c_k is $k-th$ class.
- **Key**: Distance measure
  e.g. Diffusion distance: How to define distance on the graph
  data: network $G = (V, W)$ where $V = \{x_j\}, W = \{w_{ij}\}$ represent vertices and weight of edges, respectively. For example, set weight $w_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\sigma^2}}$ e.g. Cosine measurement
- **Objective function**: Center of gravity

$$\alpha_k = \frac{1}{|c_k|} \sum_{x_j \in c_k} x_j$$

$$J_1 = \{c_1, c_2, \dots, c_K\} = \sum_{k=1}^K \sum_{x_j \in c_k} \|x_j - \alpha_k\|^2$$

$$J_2 = \frac{1}{2} \sum_{k=1}^K \frac{1}{|c_k|} \sum_{x_i, x_j \in c_k} \|x_i - x_j\|^2$$

$$Lemma : J_1 = J_2$$

- **K-means algorithm (Hard Clustering)**

- Given a clustering

$$\{x_j\} = \bigcup_{k=1}^{K} c_k$$

calculate the center of gravity of each class (at n-step)

$$\alpha_k^{(n)} = \frac{1}{|c_k^{(n)}|} \sum_{x_j \in c_k^{(n)}} x_j$$

- Reclustering (update n+1-step)

$$\forall x_j, k(j) = \arg\min_k \|x_j - \alpha_k^{(n)}\| x_j \in c_{k(j)}^{(n+1)}$$

- **Soft (probabilistic) clustering**

$$\forall x_j, p_{jk} = prob\, of \{x_j \in c_k\}$$

- **Gaussian mixture model** We have two random variables$\{x, z\}$

$$\rho_1 \sim N(\mu_1, \sigma_1), \rho_2 \sim N(\mu_2, \sigma_2)$$

$$Prob\{x \in A | z = 1\} = \int_A \rho_1 dx, Prob\{x \in A | z = 2\} = \int_A \rho_2 dx$$

$$\begin{aligned} Prob\{x \in A\} &= Prob\{x \in A | z = 1\} Prob\{z = 1\} + Prob\{x \in A | z = 2\} Prob\{z = 2\} \\ &= \pi \int_A \rho_1 dx + (1 - \pi) \int_A \rho_2 dx \end{aligned}$$

$$Prob\, density = \pi \rho_1(x) + (1 - \pi)\rho_2(x)$$

We want to solve the inverse problem: Given $\{x_j\} \sim$ mixture distribution, how to estimate the parameters of the mixture distributions? (**)

- **Likelihood** (for parameter estimation) e.g. (Single Gaussian)Estimate μ, σ

$$\{x_j\} \sim N(\mu, \sigma^2)$$

$$L(\mu, \sigma^2) = \Pi_{j=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_j-\mu)^2}{2\sigma^2}}$$

$$l(\mu, \sigma^2) = \log L(\mu, \sigma^2) = -\sum_{j=1}^{n} \frac{(x_j - \mu)^2}{2\sigma^2} - n \log \sqrt{2\pi\sigma^2}$$

Back to (**)

$$\log L(\theta) = \sum \log(\pi \rho_1(x_j) + (1 - \pi)\rho_2(x_j))$$

It is not easy to solve θ, so we use EM algorithm.

- **EM Algorithm** From step n we obtain: $\pi^{(n)}, \mu_1^{(n)}, \sigma_1^{(n)}, \mu_2^{(n)}, \sigma_2^{(n)}$
  How to proceed step $n+1$?
  - 1. E-step

$$\begin{aligned}\Delta_j &= \text{"}prob\ x_j\ comes\ from\ \rho_1\text{"}\\ &= Prob\{z=1|x_j\}\\ &= \frac{\pi^{(n)}\rho_1^{(n)}(x_j)}{\pi^{(n)}\rho_1^{(n)}(x_j)+(1-\pi^{(n)})\rho_2^{(n)}(x_j)}\end{aligned}$$

  - 2. M-step: update $\mu_1^{(n+1)}, \sigma_1^{(n+1)}, \mu_2^{(n+1)}, \sigma_2^{(n+1)}$

$$\begin{aligned}\log(L(\theta)) &= \sum_j \log(\Delta_j\rho_1(x_j)+(1-\Delta_j)\rho_2(x_j))\\ &\geq \sum_j \Delta_j\log(\rho_1(x_j))+(1-\Delta_j)\log(\rho_2(x_j))(\text{Jensen ineq.})\\ &= \text{expected}\log(L(\theta))\end{aligned}$$

$$\theta^{(n+1)} = \arg\min_\theta\{\text{expected}\log(L(\theta))\}$$

3 Dimension Reduction

- **Data**: $\{x_j\} \subset R^d$
- **Linear**
  We want:

$$F: x \subset R^d \to z \subset R^{d'}, d' << d$$
$$G = F^{-1}: z \to x$$
$$x_j \xrightarrow{F} z_j \xrightarrow{G} \tilde{x}_j$$
$$\min_{F,G}\sum_j \|x_j - \tilde{x}_j\|^2$$

  For example

$$F(x) = \beta^\top x, \beta \in R^d$$
$$G(z) = \alpha z, \alpha \in R^d$$
$$\tilde{x} = G(F(x)) = \alpha\beta^\top x$$
$$L(\alpha,\beta) = \frac{1}{2}\sum\|x_j-\tilde{x}_j\|^2 = \frac{1}{2}\sum\|x_j-\alpha\beta^\top x_j\|^2$$
$$\nabla_\alpha L = -\sum(x_j-\alpha\beta^\top x_j)\beta^\top x_j = 0$$

$$\nabla_\beta L = -\sum (x_j - \alpha\beta^\top x_j)\alpha^\top x_j = 0$$
$$\alpha = \beta$$
$$(\sum x_j x_j^\top)\beta = \sum \beta\beta^\top x_j x_j^\top \beta = \beta \sum \beta^\top x_j x_j^\top \beta = \lambda\beta$$

Why $\alpha = \beta$? Refer `http://www.deeplearningbook.org/contents/linear_algebra.html` (page 45-50)

Rewrite formula above (denote $X = \sum x_j x_j^\top$)

$$X\beta = \lambda\beta$$

The case is a special **PCA**(Principal Component Analysis). General PCA:

$$z = W_{\tilde{d}\times d} x \in R^{\tilde{d}}, \tilde{x} = V_{\tilde{d}\times d}^\top z = V^\top W x \in R^d$$
$$L(w) = \sum \|x_j - \tilde{x}_j\|^2 = \sum \|x_j - W^\top W x_j\|^2 (V = W)$$

Consider $\tilde{W} = QW$ where $Q^\top Q = I$.

$$L(\tilde{W}) = \sum \|x_j - W^\top Q^\top Q W x_j\|^2 = \sum \|x_j - W^\top W x_j\|^2 = L(W)$$
$$\nabla_W L(W) = \sum (x_j - W^\top W x_j) x_j^\top W = 0$$
$$(\sum x_j x_j^\top) W^\top = \sum W^\top W x_j x_j^\top W \to X W^\top = \Lambda W^\top$$

Where Λ is the diagonal matrix of the largest $\tilde{d}$ eigenvalues with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{\tilde{d}}$.

- **Non-Linear: Auto-encoder** Use NN represent F, G, input R^d, output $R^{\tilde{d}}$ Objective: $L(\theta) = \sum_{j=1}^n \|x_j - \tilde{x}_j\|^2$

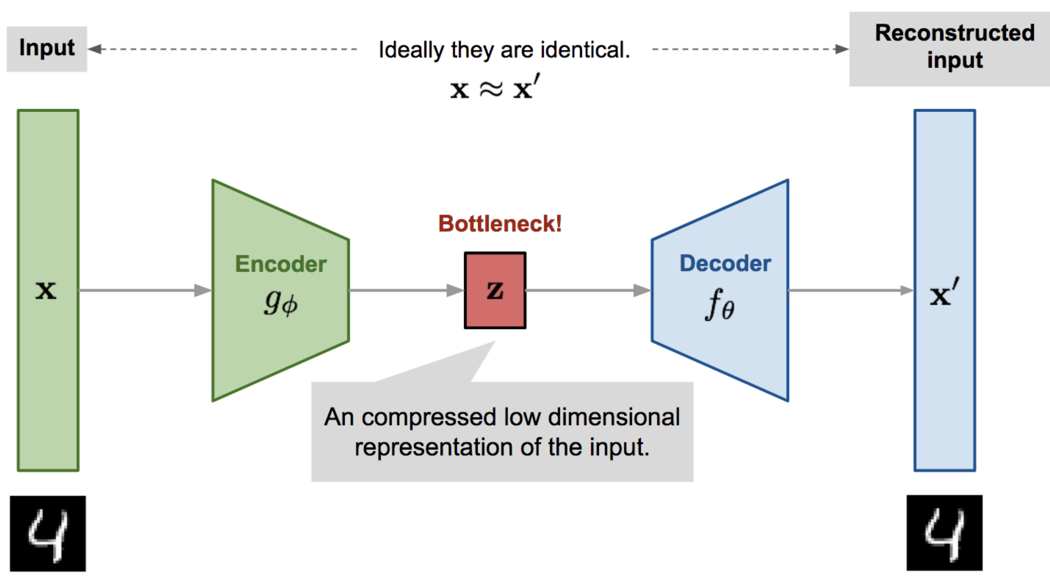


Figure Source: `https://lilianweng.github.io`

4 Density Estimation

- **Data**: $\{x_j\}$
- **Objective**: $\{x_j\} \sim \mu$ want to find μ
- **Basic idea**: histogram

$$\rho(x) \approx \frac{1}{n} \sum \frac{1}{h} H(\frac{x - x_j}{h})$$

$$\int H(x)dx = 1$$

$$H(x) = \delta(x) \text{ for example}$$

High-dimension case

$$R^{\tilde{d}} \to R^d$$

exist G, forall f

$$\int f(G(z))d\mu^* = \int f d\mu \approx \frac{1}{n} \sum_{j=1}^{n} f(x_j)$$

$$\sup_{\|f\|_{lip} \leq 1} |\int f(G(z))d\mu^* - \frac{1}{n} \sum f(x_j)| = L(\theta)$$

Where $\|\cdot\|_{lip}$ is Lipschitz norm. (Wasserstein GAN)

Macroeconomic Analysis with Machine Learning and Big Data

Lecture 4: Statistical Model in Macroeconomics (I)

Instructor: Yucheng Yang *Scribe: Yumin Hu, Lu Yang*

1 Vector Autoregressive Model

1.1 Setup of VAR

Covariance stationary process Y_t satisfies vector autoregressive model of order p (denoted as VAR(p)) if

$$Y_t = \nu + \sum_{i=1}^{p} A_i Y_{t-i} + Z_t$$

where

$$\begin{array}{lll} Y_t & \in & \mathcal{R}^{n\times 1} \\ \nu & \in & \mathcal{R}^{n\times 1} \\ A_i & \in & \mathcal{R}^{n\times n} \quad i = 1, 2, ..., p \\ Z_t & is & White\ Noise \end{array}$$

Covariance Stationary $(weakly\ stationary)$:

- $\mathbb{E}[Y_{it}^2] < \infty,\ \forall i$
- $\mathbb{E}[Y_t] \triangleq \mu_t,\ Cov(Y_t, Y_{t-l}) \triangleq \Gamma_Y(l),\ \forall l \in \mathcal{Z}\ don`t\ depend\ on\ t$

Strictly stationary:

$$(Y_t, Y_{t+1}, \cdots, Y_{t+k}) \stackrel{d}{=} (Y_{t+l}, Y_{t+l+1}, \cdots, Y_{t+k+l})$$

Z_t is White Noise if it satisfies the following conditions:

$$\mu_Z = 0,\ \Gamma_Z(l) = \begin{cases} \Sigma & l = 0 \\ 0 & l \neq 0 \end{cases}$$

Remark:

$$VAR(p) \longrightarrow VMA(\infty)$$

$$\begin{array}{rrcl} & Y_t & = & \nu + A_1 Y_{t-1} + A_2 Y_{t-2} + \cdots + A_p Y_{t-p} + Z_t \\ \Longrightarrow & Y_t & = & \nu + \sum_{i=1}^{\infty} \Psi_i Z_{t-i} \end{array}$$

Notation:

$$\begin{array}{rcl} Lag\ operator & : & \mathcal{L} \quad e.g.\ \mathcal{L}Y_t = Y_{t-1} \\ Lag\ Polynomial & : & A(\mathcal{L}) = \mathbf{I} - \sum_{i=1}^{p} A_i \mathcal{L}^i \end{array}$$

We can rewrite the model as

$$A(\mathcal{L})Y_t = Z_t.$$

Theorem: If $x \longrightarrow det(A(x))$ has all roots outside unit cycle.$\Longrightarrow \exists$ abs. sum. $A(\mathcal{L})^{-1}$, s.t. Y_t has a unique covariance stationary representation $Y_t = A(\mathcal{L})^{-1}Z_t$.
Proof. The proof of this Theorem is on Brockwell&Davis (Thm 11.3.1).

How to check whether the condition holds? For $p = 1$,

$$A(\mathcal{L}) = \mathcal{I} - A_1\mathcal{L}$$

$$det(A(x)) = 0 \Leftrightarrow |\mathbf{I} - \sum_{i=1}^{p} A_i x| = 0$$

easily calculated as the inverse of eigenvalues.

We can write VAR(p) into the companion form VAR(1) as follows:

$$X_t = (Y_t, Y_{t-1}, \cdots, Y_{t-p+1})^T$$

$$X_{t-1} = (Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p})^T$$

$$X_t = \tilde{A}X_{t-1} + \tilde{Z}_t + \tilde{\nu}$$

1.2 Estimation of VAR

For VAR(p) function

$$Y_t = \nu + \sum_{i=1}^{p} A_i Y_{t-i} + Z_t$$

Assume that $Z_t \sim \mathcal{N}(0, \Sigma)$. And all parameters $(\nu, A_1, \cdots)$ is denoted by θ.

$$P(Y_1, Y_2, \cdots, Y_T|\theta) \overset{factorization}{=} P(Y_1|\theta)\Pi_{t=2}^{T} P(Y_t|\mathcal{Y}_{t-1}, \theta)$$

Where

$$\mathcal{Y}_{t-1} = \{y_t\}_{\tau \leq t-1}$$

$$logP(Y_1, Y_2, \cdots, Y_T|\theta) = -\frac{nT}{2}log2\pi - \frac{1}{2}\sum_{t=1}^{T}[log|\hat{V}_t(\theta)| + \hat{e}_t(\theta)\hat{V}_t(\theta)^{-1}\hat{e}_t(\theta)]$$

Where

$$\begin{aligned} \hat{e}_t(\theta) &= Y_t - \mathbb{E}[Y_t|\mathcal{Y}_{t-1}, \theta] \\ \hat{V}_t(\theta) &= Var(\hat{e}_t|\theta) \end{aligned}$$

Write VAR(p) in a more compact way:

$$Y_t = \beta X_t + Z_t$$

then the CMLE estimators are:

$$\begin{aligned}\hat{\beta}_{CMLE} &= (\sum_{t=1}^{T} Y_t X_t^T)(\sum_t X_t X_t^T)^{-1} \\ \hat{\Sigma}_{CMLE} &= \frac{1}{T}\sum_{t=1}^{T}(Y_t - \hat{\beta}X_t)(Y_t - \hat{\beta}X_t)^T\end{aligned}$$

The biggest problem of VAR: small data sample size, large number of parameters - CMLE perform badly. The most common estimation methods are Bayesian methods. Nice conjugate priors for β:

$$prior : vec(\beta) \sim \mathcal{N}(\beta^*, V_\beta) \quad \Longrightarrow \quad posterior : vec(\beta) \sim \mathcal{N}(\bar{\beta}, \bar{V_\beta}).$$

e.g. Minnesota prior: prior centered at random walk.

1.3 Granger "causality"

- $\{x_t\}$ Granger cause $\{y_t\}$ if $Var^*(y_t|\mathcal{Y}_{t-1}) > Var^*(y_t|\mathcal{Y}_{t-1}, \mathcal{X}_{t-1})$
- It is just about whether it helps in prediction. Not a causality relationship.
- Sims (1972): show that money Granger cause income, not opposite.

2 Structural Vector Autoregressive Model

2.1 Setup of SVAR

- **VAR**

$$\begin{aligned}Y_t &= \nu + \sum_{i=1}^{p} A_i Y_{t-i} + \epsilon_t,\ \epsilon_t \sim N(0, \Sigma) \\ \epsilon_t &= Y_t - \mathbb{E}^*(Y_t|\mathcal{Y}_{t-1}) \text{ is one-step ahead forecasting error}\end{aligned}$$

- **SVAR**: Shocks are independent. The SVAR assumes that ϵ_t is a linear combination of the unobserved structural shocks η_t, and structural shocks are assumed to be uncorrelated:
  - $\epsilon_t = H\eta_t, \eta_t \sim \mathcal{N}(0, diag\{\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2\})$
  - VAR can be written as $Y_t = \nu + \sum_{i=1}^{p} A_i Y_{t-i} + H\eta_t$.
  - SVAR is $B_0 Y_t = C + \sum_{i=1}^{p} B_i Y_{t-i} + \eta_t,\ \eta_t \sim N(0, \Sigma_\eta)$
  - If B_0^{-1} exists $\Longrightarrow Y_t = B_0^{-1}C + \sum_{i=1}^{p} B_0^{-1} B_i Y_{t-i} + B_0^{-1}\eta_t,\ B_0^{-1}\eta_t \sim N(0, B_0^{-1}\Sigma_\eta(B_0^{-1})^T)$
  - After estimating VAR, we need to estimate B_0^{-1} to estimate SVAR.

2.2 Identification

- the estimation of B_0^{-1} requires the user to impose additional identifying restrictions on B_0^{-1} that can be motivated based on economic theory, or other external constraints on structural model. Imposing these additional restrictions allows one to decompose the reduced-form errors ϵ_t into manually uncorrelated structural shocks η_t, with an economic interpretation.

- degree of freedom: $\frac{n(n-1)}{2}$.
  - Total # of unknowns in B_0^{-1}: n^2.
  - Since $B_0^{-1}\Sigma_\eta(B_0^{-1})^T = \Sigma$ and Σ_η is diagonal: $\frac{n(n-1)}{2}$ restrictions.
  - Normalization condition ($\Sigma_\eta = \mathbf{I}$): n restrictions.

- Short-run restriction (Sims, 1980):
  - impose $\frac{n(n-1)}{2}$ zero restrictions on B_0^{-1}.
  - for example, B_0^{-1} is lower triangular matrix, then we can directly get it from Cholesky decomposition.

3 State Space Model

Observation/Measurement equation: $Y_t = g(S_t, u_t;\ \theta)$

State/Transition equation: $S_t = h(S_{t-1}, \epsilon_t;\ \theta)$

- Y_t: n-dim observable
- S_t: m-dim state(unobservable/latent)
- $(u_t^T, \epsilon_t^T) \overset{iid}{\sim} F_\theta,\ \{\epsilon_t\}$ and $\{u_t\}$ are orthogonal.

Filtering Problem: given knowledge of θ, learn S_t from data $\mathcal{Y}_t = \{Y_\tau\}_{\tau=1}^t$

- $P(S_{t-1}|\mathcal{Y}_{t-1}, \theta)$
- forecast $S:\ P(S_t|\mathcal{Y}_{t-1}, \theta) = \int P(S_t|S_{t-1}, \theta)P(S_{t-1}|\mathcal{Y}_{t-1}, \theta)dS_{t-1}$
- forecast $Y_t:\ P(Y_t|\mathcal{Y}_{t-1}, \theta) = \int P(Y_t|S_t, \theta)P(S_t|\mathcal{Y}_{t-1}, \theta)dS_t$
- update via Bayes Rule given Y_t:

$$P(S_t|\mathcal{Y}_t, \theta) = P(S_t|Y_t, \mathcal{Y}_{t-1}, \theta) = \frac{P(S_t, Y_t|\mathcal{Y}_{t-1}, \theta)}{P(Y_t|\mathcal{Y}_{t-1}, \theta)} = \frac{P(S_t|\mathcal{Y}_{t-1}, \theta)P(Y_t|S_t, \theta)}{P(Y_t|\mathcal{Y}_{t-1}, \theta)}$$

Smoothing Problem: $S_t|y_T$, given $P(S_{t+1}|y_T, \theta)$

$$
\begin{aligned}
P(S_t|\mathcal{Y}_T,\theta) &= \int P(S_t|S_{t+1},\mathcal{Y}_T,\theta)P(S_{t+1}|\mathcal{Y}_T,\theta)dS_{t+1} \\
&= \int P(S_t|S_{t+1},\mathcal{Y}_t,\theta)P(S_{t+1}|\mathcal{Y}_T,\theta)dS_{t+1} \\
&= P(S_t|\mathcal{Y}_t,\theta)\int \frac{P(S_{t+1}|S_t,\theta)}{P(S_{t+1}|\mathcal{Y}_t,\theta)}P(S_{t+1}|\mathcal{Y}_T,\theta)dS_{t+1}
\end{aligned}
$$

4 Linear Gaussian State Space Model

Special important case: Dynamic Factor Model

- Observable equation:
$$X_t = \Lambda F_t + \epsilon_t,\ \epsilon_t \in N(0,R)$$
- State equation:
$$F_t = \Phi F_{t-1} + u_t,\ u_t \in N(0,Q)$$
- How to compute the likelihood when $\{F_t\}$ is latent?
  - EM-algorithm (Expectation and Maximization)
  - PCA (Principle Component Analysis)

EM-algorithm

- E-step (Kalman filter and Kalman smoother): given $\Phi, \Lambda, Q, R \Rightarrow \{F_t\}_{t=1}^T \Big| \{X_t\}_{t=1}^T$
  - Denote:
$$
\begin{aligned}
F_t|\mathcal{X}_{t-1} &\sim N(F_{t|t-1}, P_{t|t-1}) \\
X_t|\mathcal{X}_{t-1} &\sim N(X_{t|t-1}, V_{t|t-1}) \\
F_t|\mathcal{X}_t &\sim N(F_{t|t}, P_{t|t})
\end{aligned}
$$
  - Kalman Filter: $F_t|\mathcal{X}_{t-1} \to X_t|\mathcal{X}_{t-1} \to F_t|\mathcal{X}_t \to F_{t+1}|\mathcal{X}_t \to ...$
$$
\begin{array}{rcl}
\hat{F}_{t+1|t} &=& \Phi \hat{F}_{t|t} \\
P_{t+1|t} &=& \Phi P_{t|t}\Phi^T + Q \\
\hat{F}_{t+1|t+1} &=& \hat{F}_{t+1|t} + K_{t+1}(X_{t+1} - \Lambda \hat{F}_{t+1|t}) \\
P_{t+1|t+1} &=& P_{t+1|t} - K_{t+1}\Lambda P_{t+1|t}
\end{array}
$$
    where the optimal Kalman Gain $K_{t+1} = P_{t+1|t}\Lambda^T(\Lambda P_{t+1|t}\Lambda^T + R)^{-1}$

- Kalman Smoother: $F_T|\mathcal{X}_T \to F_{T-1}|\mathcal{X}_T \to ... \to F_t|\mathcal{X}_T$

$$\hat{F}_{t|T} = \hat{F}_{t|t} + L_t(\hat{F}_{t+1|T} - \hat{F}_{t+1|t})$$

$$P_{t|T} = P_{t|t} + L_t(P_{t+1|T} - P_{t+1|t})L_t^T$$

  where $L_t = P_{t|t}\Lambda^T P_{t+1|t}^{-1}$

- M-step(update parameter): update Φ, Λ, Q, R such that expected log-likelihood is maximized.

$$\{F_t, X_t\}_{t=1}^T \Rightarrow \Phi, \Lambda, Q, R$$

- Repeat until convergence, which is guaranteed to local optimum

PCA

- $Var(F_t) = I$, columns of Λ are orthogonal
- When the condition above is satisfied, $X_t \overset{PCA}{\to} \Lambda F_t$

5 Reference

Kilian, L. and Lütkepohl, H., 2017. Structural vector autoregressive analysis. Cambridge University Press.
Webpage: `https://sites.google.com/site/lkilian2019/textbook/preliminary-chapt`

Stock, J.H. and Watson, M.W., 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In Handbook of macroeconomics (Vol. 2, pp. 415-525). Elsevier.

Jur van den Berg's notes on EM Algorithm to estimate linear Gaussian state space model: `http://arl.cs.utah.edu/resources/EM\%20Algorithm.pdf`.

Macroeconomic Analysis with Machine Learning and Big Data

Lecture 5: Statistical Model in Macroeconomics (II)

Instructor: Yucheng Yang *Scribe: Yuqing Ding, Zeru Jiang, Yiming Qiao, Sun Yang*

1 Vector Autoregressive Model with Distributional Inputs

1.1 Model Setup

- Target: add micro variables (distributional variables) into VAR and SVAR.
- Input variables:
  - Z_t : macroeconomic aggregate variables.
  - $\ell_t(x) = \log P_t(x)$: log density function for one-dimensional variable $x_{i,t}$ (could be extended to multivariate case later).
- VAR $\longrightarrow$ Functional VAR: Decompose Z_t and ℓ_t into a deterministic component $(Z_*, \ell_*(x))$ and fluctuations $Z_t = Z_* + \tilde{Z}_t, \ell_t = \ell_* + \tilde{\ell}_t$. $(\tilde{Z}_t, \tilde{\ell}_t(x))$ evolve according to linear **functional VAR**:

$$\begin{aligned} \tilde{Z}_t &= B_{zz}\tilde{Z}_{t-1} + \int B_{zl}(\tilde{x})\tilde{\ell}_{t-1}d\tilde{x} + u_{z,t} \\ \tilde{\ell}_t(x) &= B_{lz}(x)\tilde{Z}_{t-1} + \int B_{ll}(x,\tilde{x})\ell_{t-1}(\tilde{x})d\tilde{x} + u_{l,t}(x). \end{aligned} \tag{1}$$

  with kernel functions $B_{zl}(\tilde{x})$ and $B_{ll}(x,\tilde{x})$.
- Dimension reduction: how to reduce the dimension of $\tilde{\ell}_t(x)$?
  1. Moments. First moment, second moment, etc. Generalized moments also possible.
  2. Nonparametric estimation. Eg. Sieve approximation; autoencoder (popular in many real world applications).
  3. Dimension reduction with economic theory. E.g. mass of certain type of agents.
- Sieve approximations (Chang, Chen and Schorfheide, 2018):

$$\tilde{\ell}_t(x) \approx \tilde{\ell}_t^K = \sum_{k=0}^{K} \tilde{\alpha}_{k,t}\zeta_k(x) = [\zeta_0(x), \zeta_1(x), ..., \zeta_K(x)] \begin{bmatrix} \tilde{\alpha}_{0,t} \\ \vdots \\ \tilde{\alpha}_{K,t} \end{bmatrix} = \zeta_t(x)\tilde{\alpha}_t \tag{2}$$

with basis functions $\{\zeta_k(x)\}_{k=0}^{K}$. Then we can get the new VAR formulation

$$\text{fVAR Equation } (1) \Rightarrow \begin{bmatrix} \tilde{Z}_t \\ \tilde{\alpha}_t \end{bmatrix} = \begin{bmatrix} B_{zz} & B_{zl}C_\alpha \\ B_{lz} & B_{ll}C_\alpha \end{bmatrix} \begin{bmatrix} \tilde{Z}_{t-1} \\ \tilde{\alpha}_{t-1} \end{bmatrix} + \begin{bmatrix} u_{z,t} \\ u_{\alpha,t} \end{bmatrix} \tag{3}$$

Where

$$C_\alpha = \int \xi'(\tilde{x})\zeta(\tilde{x})d\tilde{x}. \tag{4}$$

with another sequence of basis functions $\{\xi_j(x)\}_{j=0}^{J}$.

- Estimation: In equation (3), we can calculate $\tilde{Z}_t, \tilde{\alpha}_t, C_\alpha$ with data and basis functions we choose. Then we may estimate Equation (3) directly using Bayesian methods, but may also estimate a state space formulation like Chang et al. (2018) if we consider measurement errors of the observed distribution.

1.2 Implementation

- Basis functions: splines. A spline is a piecewise polynomial functions with knots $x_s, s = 1, 2, ..., S$:

$$\begin{aligned} Spl(m, S) = & \sum_{k=0}^{m} a_k(x^k \mathbb{1}\{x \leq x_S\} + x_S^k \mathbb{1}\{x > x_S\}) \\ & + \sum_{s=1}^{S-1} b_s([max\{x - x_s, 0\}]^m \mathbb{1}\{x \leq x_S\} + (x_S - x_s)^m \mathbb{1}\{x > x_S\}) \\ & + \sum_{t=1}^{m} C_k[max\{x - x_S, 0\}]^m \end{aligned} \tag{5}$$

$m = 3, a_2 = a_1 = 0, c_2 = c_3 = 0 \Rightarrow$ tails of Laplace deasity
And we can obtain that

$$\begin{aligned} \zeta_0(x) &= 1 \\ \zeta_1(x) &= (x\mathbb{1}\{x \leq x_S\} + x_S\mathbb{1}\{x > x_S\}) \\ \zeta_2(x) &= ([max\{x - x_1, 0\}^3 \mathbb{1}\{x \leq x_S\} + (x_S - x_1)^3 \mathbb{1}\{x > x_S\}) \\ &\vdots \\ \zeta_S(x) &= (max\{x - x_{S-1}, 0\}^3 \mathbb{1}\{x \leq x_S\} + (x_S - x_{S-1})^3 \mathbb{1}\{x > x_S\}) \\ \zeta_{S+1}(x) &= max\{x - x_S, 0\} \end{aligned} \tag{6}$$

- Sieve Coefficients: we obtain $\hat{\alpha^0_{k,t}}$ with sieve approximation for the density function at each time t. Chang et al. (2018) also apply seasonal adjustment and compression (PCA) on $\hat{\alpha^0_{k,t}}$ to get rid of seasonality and collinearity. The new functional VAR after the seasonal adjustment and compression operations share the same structure as equation (3) which could be estimated using Bayesian method.

- All the operations so far are reversible so that we can go back to the original variables after computing impulse responses and other things with the functional VAR.
- Bayesian estimation of functional VAR like equation (3).
- Recover densities from α's.

1.3 Empirical Results

The model can be used to look into:

- VAR-type questions:
    1. How distributional dynamics affect aggregate dynamics?
    2. How aggregate dynamics affect distributional dynamics?
- SVAR-type questions (with further identification assumptions):
    1. How distributional shocks affect aggregate dynamics?
    2. How aggregate shocks affect distributional dynamics?

2 State Space Model with Distributional Inputs

Similar to vector autoregressive models, state space models can also handle micro distributional inputs. We introduce the framework proposed by Liu and Plagborg-Moller (2019) in this section.

2.1 Model Setup and Key Assumptions

The structure of the state space model is visualized in Figure 1.

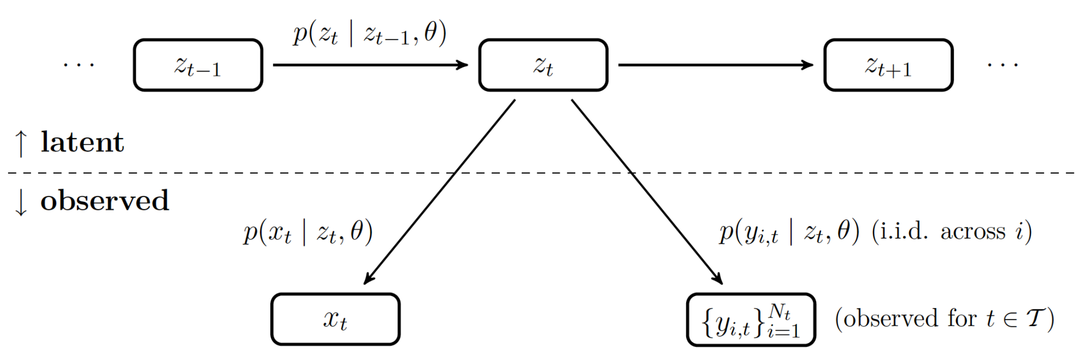


Figure 1: Diagram of state space model with both micro and macro observables.

- Observables: macro variables $x_t \sim p(x_t|z_t, \theta)$, micro distributional variables $y_{i,t} \sim p(y_{i,t}|z_t, \theta)$ and is i.i.d. across different i.
- Latent states: $z_t \sim p(z_t|z_{t-1})$. The latent variables could be observed or unobserved macro variables, could also be distributional variables (e.g. the whole distribution). However, the latent state is shared among all the individuals, and cannot be identified as $z_{i,t}$.
- Here they impose assumptions that $y_{i,t}$ is independent of $y_{i,t-1}$ given macro latent state z_t.

2.2 Likelihood and Estimation

The likelihood can be written as:

$$p(x, y|\theta) = \underbrace{p(x|\theta)}_{\text{Macro Part}} \underbrace{p(y|x, \theta)}_{\text{Distributional Part}} \tag{7}$$

The **macro part** could be obtained via standard state space likelihood formulation:

$$p(x|\theta) = \prod_{t=1}^{T} p(x_t|\mathcal{X}_{t-1}, \theta) = \prod_{t=1}^{T} p(x_t|z_t, \theta)p(z_t|z_{t-1})p(z_{t-1}|\mathcal{X}_{t-1}, \theta) \tag{8}$$

where $\mathcal{X}_t = \{x_\tau\}_{\tau=1}^{t}$ and $p(z_{t-1}|\mathcal{X}_{t-1}, \theta)$ is obtained from Kalman filter with initial assumption on $p(z_0|\mathcal{X}_0, \theta) = p(z_0|\theta)$.

The **distributional part** could be written as

$$\begin{aligned} p(y|x, \theta) &= \int p(y|z, x, \theta)p(z|x, \theta)dz = \int p(y|z, \theta)p(z|x, \theta)dz \\ &= \int \prod_{t=1}^{T} \prod_{i=1}^{N_t} p(y_{it}|z_t, \theta)p(z|x, \theta)dz \end{aligned} \tag{9}$$

If we can sample draws $\mathbf{z}^{(j)} \equiv \left\{z_t^{(j)}\right\}_{1\leq t\leq T}$, $j = 1, \ldots, J$ from the density

$$p(z|x, \theta) = \prod_{t=1}^{T} p(z_t|\mathcal{X}_t, \theta)$$

which can be obtained from the Kalman smoother, then we can approximate the distributional part of the likelihood with

$$p(y|x, \theta) \approx \frac{1}{J} \sum_{j=1}^{J} \sum_{t=1}^{T} \sum_{i=1}^{N_t} p(y_{it}|z_t = z_t^{(j)}, \theta) \tag{10}$$

3 Recurrent Neural Network and LSTM Network

One of the most important formulation of nonlinear state space models is the recurrent neural networks (RNN). We introduce RNN and its variant long short-term memory (LSTM) networks in this section.

3.1 Recurrent neural network (RNN)

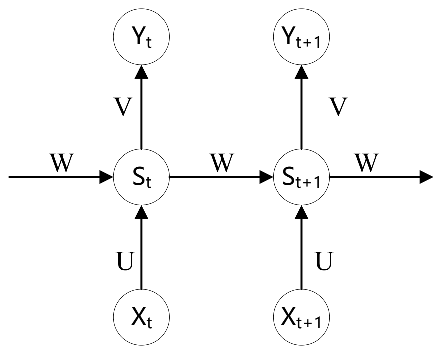


Figure 2: An unrolled recurrent neural network.

Similar to linear state space model, RNN has hidden states S_t, which affect the output Y_t together with input variables X_t as illustrated in Figure 2.

$$\begin{aligned} S_t &= \sigma(WS_{t-1} + UX_t + b) \\ Y_t &= VS_t \end{aligned}$$

where

- $X_t \in \mathcal{R}^{k\times 1}$: input variables.
- $S_t \in \mathcal{R}^{n\times 1}$: state variables.
- $Y_t \in \mathcal{R}^{m\times 1}$: output variables.
- $W \in \mathcal{R}^{n\times n}, U \in \mathcal{R}^{k\times n}, V \in \mathcal{R}^{n\times m}, b \in \mathcal{R}^{n\times 1}$: parameters.

Remark: With constant transition matrix over time, RNN suffers from so-called “gradient vanishing” problem, so it could not deal with problems with “long-term dependencies”. LSTM is proposed to handle this problem.

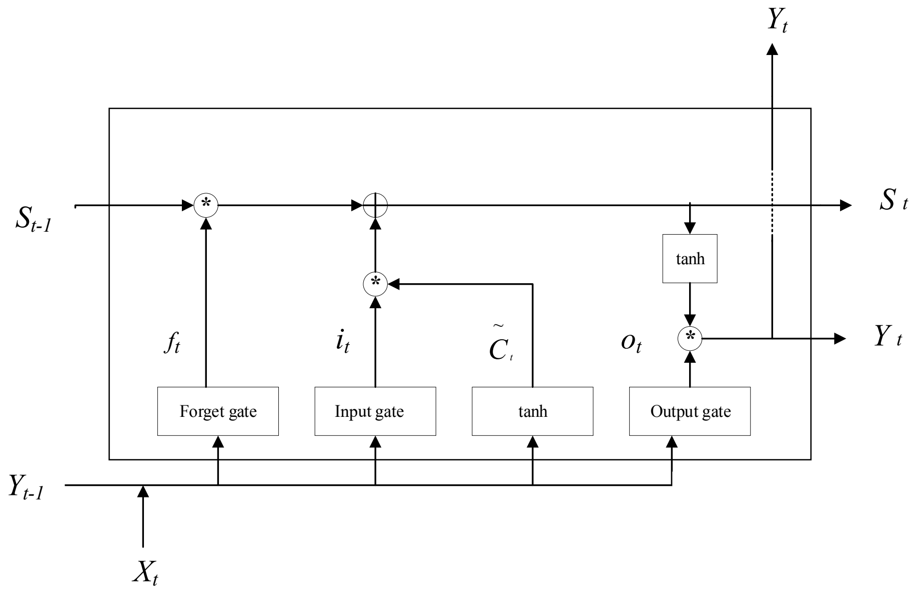


Figure 3: A LSTM cell

3.2 Long Short-term Memory (LSTM)

Figure 3 shows the detailed structure of a LSTM cell. The core of LSTM is the design of three gates: *forget gate*, *input gate* and *output gate*.

- **Forget gate.** It decides what information LSTM is going to throw away from the cell state.

$$f_t = \sigma(W_f \cdot [Y_{t-1}, X_t] + b_f)$$

  where σ indicates the sigmoid activation function:

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

- **Input gate.** It decides what new information LSTM is going to store in the cell state.

$$\begin{aligned} i_t &= \sigma(W_i \cdot [Y_{t-1}, X_t] + b_i) \\ \tilde{C}_t &= \tanh(W_c \cdot [Y_{t-1}, X_t] + b_i) \end{aligned}$$

  where $\tilde{C}_t$ is a vector of new candidates values, which will be used to update cell state.

- **Update cell state.** S_{t-1} denotes the old cell state, and then update it into new cell state S_t by:

$$S_t = f_t * S_{t-1} + i_t * \hat{S}_t$$

- **Output gate.** It decides what LSTM is going to output.

$$\begin{aligned} o_t &= \sigma(W_o \cdot [Y_{t-1}, X_t] + b_o) \\ Y_t &= o_t * \tanh(S_t) \end{aligned}$$

LSTM put the cell state through $\tanh$ ($\tan(S_t) \in [-1, 1]$) and multiply it by the output of the sigmoid gate, then get the output Y_t.

Reference

Chang, M., Chen, X. and Schorfheide, F., 2018. Heterogeneity and Aggregate Fluctuations. Manuscript, University of Pennsylvania.

Liu, L. and Plagborg-Møller, M., 2019. Full-Information Estimation of Heterogeneous Agent Models Using Macro and Micro Data.

Christopher Olah. 2015. Understanding LSTM Networks. `https://colah.github.io/posts/2015-08-Understanding-LSTMs/`.

Karpathy, A., 2015. The unreasonable effectiveness of recurrent neural networks. `http://karpathy.github.io/2015/05/21/rnn-effectiveness/`.

Macroeconomic Analysis with Machine Learning and Big Data

Lecture 6: Structural Models in Macroeconomics (I)

Instructor: Yucheng Yang *Scribe: Gefan Liu, Xilin Yuan*

1 The Development of Structural Models in Macroeconomics

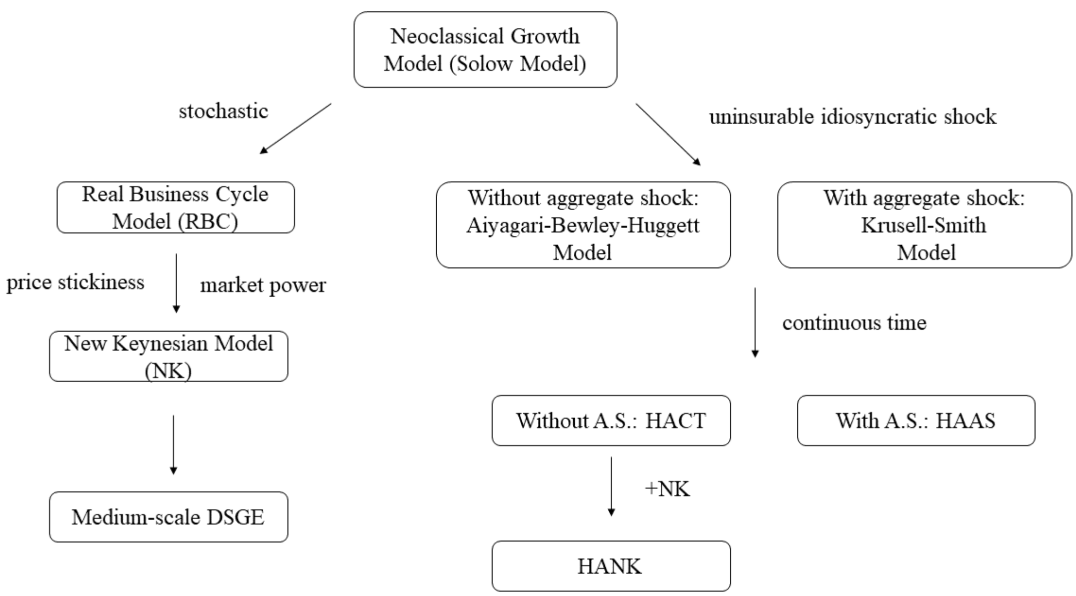


Figure 1: The Development of Structural Models in Macro

2 Real Business Cycle (RBC) Model

2.1 Planner's problem

$$\max_{\{C_t, K_{t+1}, H_t\}_t} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t [logC_t + \psi_t log(1 - H_t)]$$

$$\text{s.t.} \qquad K_{t+1} = (1-\delta)K_t + I_t$$

$$C_t + I_t = A_t K_t^{\alpha} H_t^{1-\alpha}$$

where H_t denotes labor input in time t. Assume A_t (productivity), ψ_t (preference from leisure) evolve stochastically according to AR(1) processes

$$log\psi_t = log\bar{\psi}(1-\rho) + \rho log\psi_{t-1} + \sigma_\psi \epsilon_t^\psi$$
$$logA_t = \rho logA_{t-1} + \sigma_A \epsilon_t^A$$

2.2 The Lagrangian and FOC

Lagrangian:

$$\mathscr{L} : \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \{logC_t + \psi_t log(1-H_t) + \lambda_t [A_t K_t^\alpha H_t^{1-\alpha} - C_t - K_{t+1} + (1-\delta)K_t]\}$$

F.O.C:

$$[C_t]: \quad \frac{1}{C_t} = \lambda_t$$

$$[H_t]: \quad \frac{\psi_t}{1-H_t} = \lambda_t (1-\alpha) A_t K_t^\alpha H_t^{1-\alpha}$$

$$[K_{t+1}]: \quad \lambda_t = \beta \mathbb{E}_t [\lambda_{t+1} [A_{t+1} \alpha (\frac{H_{t+1}}{K_{t+1}})^{1-\alpha} + 1 - \delta]]$$

Budget Constraint:

$$[BC]: \quad A_t K_t^\alpha H_t^{1-\alpha} = C_t + K_{t+1} - (1-\delta)K_t$$

We use $\frac{1}{C_t} = \lambda_t$ to substitute λ_t and λ_{t+1} in equations $[H_t]$ and $[K_{t+1}]$ above.

2.3 Steady state

Given initial condition such as K_0, can we get $\{C_t, K_{t+1}, H_t\}_{t=0}^{\infty}$? The answer is no, and we will focus on fluctuations around the steady state.

Definition 2.1. Steady state: $C_t \equiv C^\star, K_t \equiv K^*, H_t \equiv H^*$.

Using FOCs and BC conditions, we can solve:

1. The steady state value of hours worked:
$$H^* = [1 + \frac{\bar{\psi}}{\alpha} - \frac{\delta\beta\bar{\psi}}{1-\beta(1-\delta)}]^{-1}$$
2. The steady state capital:
$$K^* = H^* [\frac{1-\beta(1-\delta)}{\beta\alpha}]^{-\frac{1}{1-\alpha}}$$
3. The steady state consumption:
$$C^* = (K^*)^\alpha (H^*)^{1-\alpha} - \delta K^*$$

4. The steady state share of output that is invested:

$$S_i \equiv \frac{\delta K^*}{\delta K^* + C^*} = \frac{\delta\beta\alpha}{1-\beta(1-\delta)}$$

5. The steady state share of output that is consumed:

$$S_c = 1 - \frac{\delta\beta\alpha}{1-\beta(1-\delta)}$$

2.4 Log linearize

Define $\hat{x}_t = log(\frac{X_t}{X^*}) \approx \frac{X_t - X^*}{X^*}$, then we can rewrite the FOCs and BC as:

$$[H_t]: \quad \hat{\psi}_t + \frac{H^*}{1-H^*}\hat{h}_t = -\hat{c}_t + \hat{a}_t + \alpha\hat{k}_t - \alpha\hat{h}_t$$

$$[K_{t+1}]: \quad -\hat{c}_t = \mathbb{E}_t[-\hat{c}_{t+1} + (1-\beta(1-\delta))(1-\alpha)[\hat{h}_{t+1} - \hat{k}_{t+1}]] + (1-\beta(1-\delta))\mathbb{E}_t\hat{a}_{t+1}$$

$$[BC]: \quad \hat{a}_t + \alpha\hat{k}_t + (1-\alpha)\hat{h}_t = S_c\hat{c}_t + \frac{S_i}{\delta}[\hat{k}_{t+1} - (1-\delta)\hat{k}_t$$

More details on log-linearization can be found in Eric Sims' notes `https://www3.nd.edu/~esims1/log_linearization_sp17.pdf`

2.5 Solving Linear Rational Expectations Models

Plug Equation $[H_t]$ into the other two log-linearized equations $\rightarrow$ 2 x 2 system of linear equations. Matrix form:

$$\mathbf{A}\begin{pmatrix}\mathbb{E}_t[\hat{c}_{t+1}] \\ \hat{k}_{t+1}\end{pmatrix} = \mathbf{B}\begin{pmatrix}\hat{c}_t \\ \hat{k}_t\end{pmatrix} + \mathbf{C}\begin{pmatrix}\hat{a}_t \\ \hat{\psi}_t\end{pmatrix} + \mathbf{Z}\mathbb{E}_t\begin{pmatrix}\hat{a}_{t+1} \\ \hat{\psi}_{t+1}\end{pmatrix}$$

$$\begin{pmatrix}\hat{a}_{t+1} \\ \hat{\psi}_{t+1}\end{pmatrix} = \rho\begin{pmatrix}\hat{a}_t \\ \hat{\psi}_t\end{pmatrix} + \begin{pmatrix}\sigma^A\epsilon_t^A \\ \sigma^\psi\epsilon_t^\psi\end{pmatrix}$$

A general solution method is discussed in Sims (2003) [1]. Here we only present a special case. Define $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{G} = [\mathbf{A}^{-1} + \rho\mathbf{Z}]\mathbf{C}$, then

$$\begin{pmatrix}\mathbb{E}_t[\hat{c}_{t+1}] \\ \hat{k}_{t+1}\end{pmatrix} = \mathbf{F}\begin{pmatrix}\hat{c}_t \\ \hat{k}_t\end{pmatrix} + \mathbf{G}\begin{pmatrix}\hat{a}_t \\ \hat{\psi}_t\end{pmatrix} \tag{1}$$

Next Goal: Get rid of $\hat{c}_{t+1}$ and write the linearized policy function

$$\hat{k}_{t+1} = M_{kk}\hat{k}_t + M_{ka}\hat{a}_t + M_{k\psi}\hat{\psi}_t$$

$$\hat{c}_t = M_{ck}\hat{k}_t + M_{ca}\hat{a}_t + M_{c\psi}\hat{\psi}_t$$

[1] `http://sims.princeton.edu/yftp/gensys/LINRE3A.pdf`

If F is diagonalizable, we may write

$$F = VDV^{-1}$$

where D is diagonal. Suppose, without loss of generality $D_1 > D_2$. Multiply both sides by V^{-1},we can get

$$V^{-1}\begin{pmatrix}\mathbb{E}_t[\hat{c}_{t+1}]\\ \hat{k}_{t+1}\end{pmatrix} = DV^{-1}\begin{pmatrix}\hat{c}_t\\ \hat{k}_t\end{pmatrix} + V^{-1}G\begin{pmatrix}\hat{a}_t\\ \hat{\psi}_t\end{pmatrix}$$
$$\mathbb{E}_t\begin{bmatrix}\Upsilon_{1,t+1}\\ \Upsilon_{2,t+1}\end{bmatrix} = \begin{pmatrix}D_1 & 0\\ 0 & D_2\end{pmatrix}\begin{bmatrix}\Upsilon_{1,t}\\ \Upsilon_{2,t}\end{bmatrix} + V^{-1}G\begin{pmatrix}\hat{a}_t\\ \hat{\psi}_t\end{pmatrix}$$

Take the first equation

$$\mathbb{E}_t\Upsilon_{1,t+1} = D_1\Upsilon_{1,t} + (V^{-1}G)_{1\cdot}\begin{pmatrix}\hat{a}_t\\ \hat{\psi}_t\end{pmatrix}$$

Rewrite VAR(1) as VMA: for τ periods head

$$\mathbb{E}_t\Upsilon_{1,t+\tau} = (D_1)^{\tau}\Upsilon_{1,t} + \sum_{i=0}^{\tau-1} D_1^{\tau-i-1}(V^{-1}G)_{1\cdot}\mathbb{E}_t\begin{pmatrix}\hat{a}_{t+i}\\ \hat{\psi}_{t+i}\end{pmatrix}$$

Re-arranging this equation:

$$\begin{aligned}\Upsilon_{1,t} &= (D_1)^{-\tau}\mathbb{E}_t\Upsilon_{1,t+\tau} + \sum_{i=0}^{\tau-1} D_1^{-i-1}(V^{-1}G)_{1\cdot}\mathbb{E}_t\begin{pmatrix}\hat{a}_{t+i}\\ \hat{\psi}_{t+i}\end{pmatrix}\\ &= \sum_{i=0}^{\infty} D_1^{-i-1}(V^{-1}G)_{1\cdot}\mathbb{E}_t\begin{pmatrix}\hat{a}_{t+i}\\ \hat{\psi}_{t+i}\end{pmatrix}\\ &= D_1^{-1}(I - D_1^{-1}\rho)^{-1}(V^{-1}G)_{1\cdot}\begin{pmatrix}\hat{a}_t\\ \hat{\psi}_t\end{pmatrix}\end{aligned}$$

Need the larger eigenvalue to be bigger than 1 for the sum to converge. By definition of Υ:

$$\Upsilon_{1,t} = (V^{-1})_{11}\hat{c}_t + (V^{-1})_{12}\hat{k}_t = D_1^{-1}(I - D_1^{-1}\rho)^{-1}(V^{-1}G)_{1\cdot}\begin{pmatrix}\hat{a}_t\\ \hat{\psi}_t\end{pmatrix}$$

Now we can write the policy variable $\hat{c}_t$ as a function of $\hat{k}_t$ and $\begin{pmatrix}\hat{a}_t\\ \hat{\psi}_t\end{pmatrix}$. We can also replace the $\hat{c}_t$ in the second line of equation (1) so that we can write $\hat{k}_{t+1}$ as a function of $\hat{k}_t$ and $\begin{pmatrix}\hat{a}_t\\ \hat{\psi}_t\end{pmatrix}$.

3 Recursive Method

3.1 Model Setup

A simple example of a sequential problem that can be solved by recursive method is

$$
\begin{aligned}
&\max_{\{k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t) \\
&s.t.\ c_t + k_t = f(k_t) \\
&\quad 0 \leq k_{t+1} \leq f(k_t),\ k_0\ given
\end{aligned}
\tag{2}
$$

This can be written as

$$
\max_{\{k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t) = \max_{\{k_1\}} u(f(k_0) - k_1) + \beta \max_{\{k_{t+1}\}_{t=1}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(f(k_{t+1}) - k_{t+2}) \tag{3}
$$

Motivated by this form, we can formulate the Bellman equation

$$
V(k) = \max_{k'} \{u(f(k) - k') + \beta V(k')\} \tag{4}
$$

$V(\cdot)$ is the value function of the Bellman equation, k is the state variable, and k' is the control variable. Under certain conditions, the value function of the sequential problem (the total discounted utility) (2) is equivalent to the solution to the functional equation (4).

3.2 Value Function Iteration

The functional equation 4 can be solved with various methods. One method is iteration over the value function V by the following algorithm.

- Make a initial guess of $V_0(k)$
- For a given $V_n(k)$, solve the maximization problem

$$
V_{n+1}(k) = \max_{k'} \{u(f(k) - k') + \beta V_n(k')\} \tag{5}
$$

- Compare V_n and V_{n+1}. If the error falls in the range of tolerance, then V_{n+1} is the solution. Otherwise, go back to step 2 and repeat until convergence.

4 Aiyagari-Bewley-Huggett Model

4.1 Model Setup

In this model, we only consider the household side, whose income is exogenous and follows a finite state Markov chain. Denote all the possible incomes by $Y = \{y_1, y_2, \cdots y_n\}$. The transitional probability matrix $\Pi = (\pi_{ij})_{N \times N}$, i.e.

$$
\mathbb{P}(y_{t+1} = y_j | y_t = i) = \pi_{ij} \tag{6}
$$

In addition, In each period t, the household can save and borrow an amount of a_{t+1} at an endogenous interest rate r_t. When borrowing, the household faces a debt constraint $a_{t+1} \geq -b$, where $-b$ is a given lower limit. For the whole economy at period t, the distribution of households with income y and asset a is denoted by $\Phi_t(a, y)$.
Formally, for a household, the optimization problem can be stated as

$$\begin{aligned} \max\ \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \\ s.t.\ c_t + a_{t+1} = y_t + (1 + r_t)a_t \\ a_{t+1} \geq -b \end{aligned} \tag{7}$$

The corresponding Bellman equation is

$$\begin{aligned} V(a, y, \Phi) = \max_{a'} \{u(c) + \beta \sum_{y' \in Y} \pi(y'|y) V(a', y', \Phi')\} \\ s.t.\ c + a' = y + (1 + r(\Phi))a \\ a' \geq -b \\ \Phi' = H(\Phi) \end{aligned} \tag{8}$$

The law of motion H of the distribution Φ_t is

$$\Phi'(\mathcal{A}, \mathcal{Y}) = \int Q((a, y), (\mathcal{A}, \mathcal{Y}))\Phi(da, dy) \tag{9}$$

The transitional probability function Q satisfies

$$Q((a, y), (\mathcal{A}, \mathcal{Y})) = \sum_{y' \in \mathcal{Y}} \pi(y'|y) \mathbb{I}_{\{a'(a,y) \in \mathcal{A}\}} \tag{10}$$

where $\mathbb{I}$ is an indicator function.

4.2 Definition of Recursive Competitive Equilibrium

Definition The recursive competition equilibrium (RCE) in this model is:

- Given the interest rate r_t. The policy function $a'(a, y, \Phi)$ solves the household optimization problem (8), with the corresponding value function being $V(a, y, \Phi)$.
- Market clearing, i.e., the amount of net saving is 0 and goods market clear for all Φ.

$$\int a'(a, y, \Phi) d\Phi = 0 \tag{11}$$

$$\int c(a, y) d\Phi = \int y d\Phi \tag{12}$$

- Law of motion H for Φ_t is generated by the exogenous Markov process π and the policy function a' as in Equations 9 and 10.

However, since the distribution Φ_t is time-varying, it is extremely difficult to solve the model even numerically. To address this problem, we further refine our definition of equilibrium to **stationary recursive competitive equilibrium** by imposing the following assumption.

- Φ is a stationary distribution. That is

$$\Phi_{t+1} = \Phi_t = \Phi^* \tag{13}$$

Given Φ is unchanged, the interest rate r_t is also time-invariant. Assume $r_t = r^*$

4.3 Solving Stationary Equilibrium by Recursive Method

We can solve the stationary equilibrium through the following algorithm

1. Make a initial guess of r. Then, solve the Bellman equation of the household problem with standard methods like value function iteration and obtain V_r, a'_r, and c_r.

2. With the policy function a' and the transitional probability matrix Π, we can get the law of motion H_r for the asset and income joint distribution, and then compute the stationary distribution Φ^*_r.

3. With Φ^*_r, a'_r, we can now compute the net saving. If the market clearing condition 11 is satisfied, then the model is solved. If the net saving is great than 0, then we should lower the guess of r, and *vice versa*. Go back to step 1 and repeat until convergence.

Lecture 7: HACT and Mean Field Games (MFG)

Instructor: Jiequn Han *Scribe: Mian Qin, Yunjie Song, Yucheng Yang*

Three advantages of heterogeneous agent model: (1) more realistic micro foundation, (2) better match the micro data, (3) useful for welfare analysis for distributional issues.

1 Recap: Huggett Model (Discrete Time)

- **Model Setup**

$$\min_{c_t \geq 0, a_{t+1} \geq \underline{a}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u\left(c_t\right)$$

$$\text{s.t. } c_t + a_{t+1} = y_t + (1 + r_t)a_t$$

with a_0, y_0 given and y_t follows Markov process with transition matrix Π on possible set Y.

- **Market Clearing Condition**

$$\int_{\underline{a}}^{\infty} a'(a, y, \Phi) d\Phi = B$$

where Φ is the distribution over $\mathcal{A} \times \mathcal{Y}$, $\mathcal{A}$ is $\sigma-$algebra over possible assets, $\mathcal{Y}$ is $\sigma-$algebra over possible set Y. B is fixed supply of asset (bond) in the economy.

- **Recursive Formulation: Bellman Equation**

$$V(a, y, \Phi) = \max_{c \geq 0, a' \geq \underline{a}} u(c) + \beta \sum_{y' \in Y} \pi(y' \mid y) V(a', y', \Phi')$$

$$c + a' = y + (1 + r)a$$

$$\Phi' = H(\Phi)$$

where $H(\cdot)$ is law of motion for the joint distribution Φ, which is determined by the transition matrix Π and optimal policy function $a'(a, y, \Phi)$. Recall from the last lecture, we can only solve for stationary equilibrium where $\Phi' = \Phi$ with recursive guesses on r.

However, if we formulate this problem in continuous time, we can take advantage of advanced mathematical tools of PDEs and bring both new insights and computational convenience to heterogeneous agent modelling.

2 Heterogeneous Agent Model in Continuous Time (HACT)

2.1 Model Setup

$$\max_{c_t \geq 0, a_t \geq \underline{a}} \mathbb{E}_0 \int_0^\infty e^{-\rho t} u(c_t)\, dt$$
$$\dot{a}_t = y_t + r_t a_t - c_t$$

with a_0, y_0 given. We assume that income follows a two-state Poisson process $y_t \in \{y_1, y_2\}$ with $y_1 < y_2$. y_t jumps from state 1 to state 2 with intensity λ_1 and vice versa with intensity λ_2. The market clearing condition is

$$\int_{\underline{a}}^\infty a g_1(a,t) da + \int_{\underline{a}}^\infty a g_2(a,t) da = B$$

where $g_j(a,t)$ is the pdf over a, y_j at time t.

2.2 Recursive Formulation: Hamilton-Jacobi-Bellman (HJB) equation and Kolmogorov Forward (KF) equation

In HACT, the problem can be written as a Hamilton-Jacobi-Bellman (HJB) equation for optimization coupled with a Kolmogorov Forward (KF) equation for law of motions. We derive both equations for transition dynamics and obtain the stationary formulation for free.

Deriving HJB equation Consider the problem in discrete time with short periods of length Δt. Individuals discount the future with discount factor $\beta(\Delta t) = e^{-\rho \Delta t}$. Individuals with income y_j keep their income with probability $p_j(\Delta t) = e^{-\lambda_j \Delta t}$ and switch to y_{-j} with probability $1 - p_j(\Delta t)$. As $\Delta t \to 0$,

$$\beta(\Delta t) = e^{-\rho \Delta t} \approx 1 - \rho \Delta t, \quad p_j(\Delta t) = e^{-\lambda_j \Delta t} \approx 1 - \lambda_j \Delta t$$

Therefore, the value function reads

$$v_j(a,t) = \max_c u(c)\Delta t + (1 - \rho \Delta t)[(1 - \lambda_j \Delta t) v_j(a_{t+\Delta t}, t + \Delta t) + (\lambda_j \Delta t) v_{-j}(a_{t+\Delta t}, t+\Delta t)]$$

Rearranging the equation,

$$0 = [v_j(a_{t+\Delta t}, t+\Delta t) - v_j(a,t)] + \max_c \Delta t u(c) + \Delta t \lambda_j v_{-j}(a_{t+\Delta t}, t+\Delta t)$$
$$-\Delta t \rho v_j(a_{t+\Delta t}, t+\Delta t) - \Delta t \lambda_j v_j(a_{t+\Delta t}, t+\Delta t) + O(\Delta t^2)$$

Dividing by Δt, taking $\Delta t \to 0$, and using that

$$\lim_{\Delta t \to 0} \frac{v_j(a_{t+\Delta t}, t+\Delta t) - v_j(a,t)}{\Delta t} = \partial_a v_j(a,t)(y_j + r_t a - c) + \partial_t v_j(a,t)$$

we obtain the following HJB equation:

$$\rho v_j(a,t) = \max_c u(c) + \partial_a v_j(a,t)(y_j + r_t a - c) + \lambda_j (v_{-j}(a,t) - v_j(a,t)) + \partial_t v_j(a,t)$$

Deriving KF equation

- First we denote $\tilde{a}_t, \tilde{y}_t$ as random variables, with a_t, y_t as their possible realizations.
- Saving policy function $s_j(a_t, t) = y_j + r_t a_t - c_j(a_t, t)$.
- Wealth evolves as $d\tilde{a}_t = s_j(\tilde{a}_t, t)dt$.
- We define density $g_j(a,t), j = 1, 2$ before, but when deriving KF equation it's easier to work with the cdf $G_j(a,t) = \Pr(\tilde{a}_t \leq a, \tilde{y}_t = y_j)$.

Consider the discrete-time situation[1]:

- Step 1: Individuals make their saving decisions according to $\tilde{a}_t = \tilde{a}_{t+\Delta t} - s_j(\tilde{a}_t)\Delta t$;
- Step 2: After saving decisions are made, next period's income $\tilde{y}_{t+\Delta t}$ is realized, individuals switches from y_j to y_{-j} with probability $\lambda_j \Delta t$. So we have:

$$\begin{aligned}\Pr(\tilde{a}_{t+\Delta t} \leq a, \tilde{y}_{t+\Delta t} = y_j) = (1 - \Delta t \lambda_j) &\Pr(\tilde{a}_t \leq a - \Delta t s_j(a,t), \tilde{y}_t = y_j) \\ &+ \Delta t \lambda_{-j} \Pr(\tilde{a}_t \leq a - \Delta t s_{-j}(a), \tilde{y}_t = y_{-j})\end{aligned}$$

In terms of $G_j(\cdot, \cdot)$, this can be written as:

$$G_j(a, t + \Delta t) = (1 - \Delta t \lambda_j) G_j(a - \Delta t s_j(a,t), t) + \Delta t \lambda_{-j} G_{-j}(a - \Delta t s_{-j}(a), t)$$

Subtracting $G_j(a,t)$ from both sides and dividing by Δt,

$$\begin{aligned}\frac{G_j(a, t + \Delta t) - G_j(a,t)}{\Delta t} =& \frac{G_j(a - \Delta t s_j(a,t), t) - G_j(a,t)}{\Delta t} - \lambda_j G_j(a - \Delta t s_j(a,t), t) \\ &+ \lambda_{-j} G_{-j}(a - \Delta t s_{-j}(a), t)\end{aligned}$$

Taking the limit as $\Delta t \to 0$ gives $\partial_t G_j(a,t) = -s_j(a,t)\partial_a G_j(a,t) - \lambda_j G_j(a,t) + \lambda_{-j} G_{-j}(a,t)$
Take the derivative with respect to a, we obtain KF equation

$$\partial_t g_j(a,t) = -\partial_a [s_j(a,t) g_j(a,t)] - \lambda_j g_j(a,t) + \lambda_{-j} g_{-j}(a,t)$$

Collections of HJB and KF equations for transition dynamics

$$\rho v_j(a,t) = \max_c u(c) + \partial_a v_j(a,t)(y_j + r_t a - c) + \lambda_j (v_{-j}(a,t) - v_j(a,t)) + \partial_t v_j(a,t)$$

$$\partial_t g_j(a,t) = -\partial_a [s_j(a,t) g_j(a,t)] - \lambda_j g_j(a,t) + \lambda_{-j} g_{-j}(a,t)$$

Actually both equations could be derived directly using the infinitesimal generator of the Poisson process.

[1] To note, the KF equation would be the same if the individual realize income first then make saving decisions.

HJB and KF Equations in the Stationary Equilibrium

$$\rho v_j(a) = \max_c u(c) + v_j'(a)\left(y_j + ra - c\right) + \lambda_j\left(v_{-j}(a) - v_j(a)\right)$$

$$0 = -\frac{d}{da}\left[s_j(a)g_j(a)\right] - \lambda_j g_j(a) + \lambda_{-j} g_{-j}(a)$$

We then obtain the following Euler equation with envelope condition and FOC:

$$(\rho - r)u'\left(c_j(a)\right) = u''\left(c_j(a)\right)c_j'(a)s_j(a) + \lambda_j\left(u'\left(c_{-j}(a)\right) - u'\left(c_j(a)\right)\right)$$

Informally using Ito's formula it can be written as

$$\frac{\mathbb{E}_t\left[du'\left(c_j\left(a_t\right)\right)\right]}{u'\left(c_j\left(a_t\right)\right)} = (\rho - r)dt$$

2.3 Borrowing Constraint and Boundary Conditions

In continuous time formulation, the borrowing constraint $a_t \geq \underline{a}$ never binds in the interior of the state space, so FOC $u'(c_j(a)) = v_j'(a)$ always holds for $a > \underline{a}$. The borrowing constraint only requires a boundary condition at $\underline{a}$, which makes the problem much easier to solve. For the stationary equilibrium, the condition is $v_j'(\underline{a}) \geq u_j'(y_j + r\underline{a})$. For transition dynamics, it is $\partial_a v_j(\underline{a}, t) \geq u_j'(y_j + r_t\underline{a})$. The transition dynamics problem also requires initial condition for $g : g_j(a, 0) = g_{j,0}(a)$ and terminal condition for $v : v_j(a, T) = v_{j,\infty}(a)$.

2.4 Analytical Results of HACT

2.4.1 Consumption and Saving Behavior of the Poor

Assumption 1 The coefficient of absolute risk aversion $R(c) := -u''(c)/u'(c)$ when wealth a approaches the borrowing limit $\underline{a}$ is finite, that is

$$\underline{R} := -\lim_{a \to \underline{a}} \frac{u''\left(y_1 + ra\right)}{u'\left(y_1 + ra\right)} < \infty$$

Proposition 1 (MPCs and Saving at Borrowing Constraint) Assume that $r < \rho$, $y_1 < y_2$ and that Assumption 1 holds. Then the solution to the HJB equation and the corresponding policy functions have the following properties:
1) $s_1(\underline{a}) = 0$ but $s_1(a) < 0$ for all $a > \underline{a}$. That is, only individuals exactly at the borrowing constraint are constrained, whereas those with wealth $a > \underline{a}$ are unconstrained and decumulate assets.
2) as $a \to \underline{a}$, the saving and consumption policy function of the low income type and the corresponding instantaneous marginal propensity to consume satisfy

$$s_1(a) \sim -\sqrt{2\nu_1}\sqrt{a - \underline{a}} \qquad (19)$$

$$c_1(a) \sim y_1 + ra + \sqrt{2\nu_1}\sqrt{a - \underline{a}}$$

$$c_1'(a) \sim r + \sqrt{\frac{\nu_1}{2(a - \underline{a})}} \qquad (20)$$

$$\nu_1 := \frac{(\rho - r)u'(\underline{c}_1) + \lambda_1 (u'(\underline{c}_1) - u'(\underline{c}_2))}{-u''(\underline{c}_1)}$$

$$\approx (\rho - r)\,\mathrm{IES}(\underline{c}_1)\,\underline{c}_1 + \lambda_1(\underline{c}_2 - \underline{c}_1) \qquad (21)$$

2.4.2 Consumption and Saving Behavior of the Wealthy

Proposition 2 Assume that $r < \rho$, $y_1 < y_2$ and that relative risk aversion $-cu''(c)/u'(c)$ is bounded above for all c.
1) Then there exists $a_{\max} < \infty$ such that $s_j(a) < 0$ for all $a \geq a_{\max}, j = 1, 2$, and $s_2(a) \sim \zeta_2 (a_{\max} - a)$ as $a \to a_{\max}$ for some constant ζ_2. The wealth of an individual with initial wealth a_0 and successive high income draws y_2 converges to $a_{\max}$ asymptotically (i.e. not in finite time): $a(t) - a_{\max} \sim e^{-\zeta_2 t}(a_0 - a_{\max})$
2) In the special case of CRRA utility $u(c) = \frac{c^{1-\gamma}}{1-\gamma}$ individual policy functions are asymptotically linear in a. As $a \to \infty$, they satisfy

$$s_j(a) \sim \frac{r - \rho}{\gamma}a, \; c_j(a) \sim \frac{\rho - (1-\gamma)r}{\gamma}a \qquad (28)$$

2.5 Numerical Method to Solve the HJB and KF Equations

We present the numerical method to solve the stationary equilibrium with a special case with no income uncertainty $y_1 = y_2 = y$ in this section. The generalization to more complicated cases is straightforward. Similar to the recursive method to solve a discrete-time Huggett model, we first guess a value for r in the stationary equilibrium, then solve the HJB and KF equations respectively, then update r with the market clearing condition and go into the next iteration.

2.5.1 Finite Difference Method for HJB

In the special case with no income uncertainty $y_1 = y_2 = y$, the HJB equation is:

$$\rho v(a) = \max_c u(c) + v'(a)(y + ra - c) \qquad (51)$$

The finite difference method approximates the function v at I discrete points in the space dimension, $a_i, i = 1, \ldots, I$. We use equispaced grids, with distance Δa between grid points, and denote $v_i = v(a_i), v_i' = v'(a_i), c_i = c(a_i), s_i = s(a_i)$. The derivative v_i' is approximated with either a forward or a backward difference approximation

$$v_i' \approx \frac{v_{i+1} - v_i}{\Delta a} =: v_{i,F}' \quad \text{or} \quad v_i' \approx \frac{v_i - v_{i-1}}{\Delta a} =: v_{i,B}'$$

The choice of forward or backward difference approximations is critical to the numerical performance and will be determined via the upwind scheme discussed below. The finite difference approximation to (51) is

$$\rho v_i = u\left(c_i\right) + v_i' s_i \qquad (52)$$

with

$$s_i := y + r a_i - c_i, \quad c_i = \left(u'\right)^{-1}\left(v_i'\right), \quad i = 1, \ldots, I$$

Upwind Scheme The choice of forward or backward difference approximations to v_i' is determined via the so-called "upwind scheme". The rough idea is to use a forward difference approximation whenever the drift of the state variable (here, saving $s_i = y + r a_i - c_i$) is positive and to use a backward difference whenever it is negative. To note, in the Kolmogorov Forward equation, the choice is opposite. The upwind version of (52) is then

$$\rho v_i = u\left(c_i\right) + \frac{v_{i+1} - v_i}{\Delta a} s_i^{+} + \frac{v_i - v_{i-1}}{\Delta a} s_i^{-}, \quad i = 1, \ldots, I \qquad (53)$$

where $s_i^{+} = \max\left\{s_i, 0\right\}$ and $s_i^{-} = \min\left\{s_i, 0\right\}$.

The problem is actually more complicated as calculating s_i requires the value of v_i' as $s_i = y + r a_i - c_i$ with $c_i = \left(u'\right)^{-1}\left(v_i'\right)$. So we have to adopt an iterative algorithm to solve the HJB: with initial guess of $v_i^0, i = 1, ..., I$, for $n = 0, 1, ...$:

- Step 1: compute $(v_{i,F}^n)'$, $(v_{i,B}^n)'$, and compute $c_{i,F}^n, s_{i,F}^n$ with $(v_{i,F}^n)'$, compute $c_{i,B}^n, s_{i,B}^n$ with $(v_{i,B}^n)'$. Then compute

  $$(v_i^n)' = (v_{i,F}^n)' \mathbf{1}_{s_{i,F}^n > 0} + (v_{i,B}^n)' \mathbf{1}_{s_{i,B}^n < 0} + (\bar{v}_i^n)' \mathbf{1}_{s_{i,F}^n \leq 0 \leq s_{i,F}^n}$$

  where $(\bar{v}_i^n)' = u'(y + r a_i)$ is v_i' with zero savings.

- Step 2: with $(v_i^n)'$, compute c_i^n.

- Step 3: update v_i^{n+1} with explicit method or implicit method until v_i^{n+1} is close enough to v_i^n. The implicit method goes as:

  $$\frac{v_i^{n+1} - v_i^n}{\Delta} + \rho v_i^{n+1} = u(c_i^n) + \frac{v_{i+1}^{n+1} - v_i^{n+1}}{\Delta a} (s_{i,F}^n)^{+} + \frac{v_i^{n+1} - v_{i-1}^{n+1}}{\Delta a} (s_{i,B}^n)^{-}$$

  It can be written as a matrix form

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

  where $\mathbf{A}$ is a tridiagonal Poission transition matrix with all diagonal entries are negative, all off-diagonal entries are positive and all rows sum to zero.

Boundary Condition Remember we turn the borrowing constraint into boundary condition at $\underline{a}$. For numerics, it is

$$v'_{1,B} = u'(y + ra_1)$$

only for the backward difference approximation. The forward difference remains $v'_{1,F} = (v_2 - v_1)/\Delta a$. From (53) we can see that the boundary condition is only imposed if $s_1 < 0$.

2.5.2 Solving the KF Equation

Given v_i's and s_i's, we can solve the KF equation $0 = -\frac{d}{da}[g(a)s(a)]$ directly. It can be discretized as $0 = -[g_i s_i]'$, and the upwind scheme writes

$$0 = -\frac{[(s^n_{i,F})^+ g_i - (s^n_{i-1,F})^+ g_{i-1}]}{\Delta a} - \frac{[(s^n_{i+1,B})^- g_{i+1} - (s^n_{i,B})^- g_i]}{\Delta a}.$$

In the matrix form, it is

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

where $\mathbf{A}^T$ is the transpose of the matrix A in the HJB equation.

3 Extensions of the Models

3.1 Diffusion Income Processes

Our baseline model assumed that income y_t takes one of two values, high and low. We now extend it to a diffusion process. Assume that individual income evolves stochastically over time on a bounded interval $[\underline{y}, \overline{y}]$ with $\overline{y} > \underline{y} \geq 0$, according to the stationary diffusion process:

$$dy_t = \mu(y_t)\,dt + \sigma(y_t)\,dW_t \qquad (54)$$

W_t is a Wiener process or standard Brownian motion and the functions μ and σ are called the drift and the diffusion of the process. We normalize the process such that its stationary mean equals one.

Similar to Section 1, a *stationary* equilibrium can be written as a system of partial differential equations. The problem of individuals and the joint distribution of income and wealth satisfy stationary HJB and KF equations:

$$\rho v(a,y) = \max_c u(c) + \partial_a v(a,y)(y + ra - c) + \partial_y v(a,y)\mu(y) + \frac{1}{2}\partial_{yy} v(a,y)\sigma^2(y) \qquad (55)$$

$$0 = -\partial_a(s(a,y)g(a,y)) - \partial_y(\mu(y)g(a,y)) + \frac{1}{2}\partial_{yy}\left(\sigma^2(y)g(a,y)\right) \qquad (56)$$

Importantly, the computational algorithm laid out in Section 3 carries over without change: from a computational perspective it is immaterial whether we solve a system of ODEs like (7) and (8) or a system of PDEs like (55) and (56).

3.2 An Alternative Way of Closing the Model: Aiyagari (1994)

Section 1 assumed that wealth takes the form of bonds that are in fixed supply. We can assume as in Aiyagari (1994) that wealth takes the form of productive capital that is used by a representative firm which also hires labor. Each individual's income is the product of an economy-wide wage w_t and her idiosyncratic labor productivity z_t and her wealth follows (2) with $y_t = w_t z_t$. The total amount of capital supplied in the economy equals the total amount of wealth. In a stationary equilibrium it is given by:

$$K = \int_{\underline{z}}^{\overline{z}} \int_{\underline{a}}^{\infty} a g(a,z) da dz := S(r,w) \qquad (60)$$

Capital depreciates at rate δ. There is a representative firm with a constant returns to scale production function $Y = F(K,L)$. Since factor markets are competitive, the wage and the interest rate are given by:

$$r = \partial_K F(K,1) - \delta, \quad w = \partial_L F(K,1) \qquad (61)$$

where we use that the mean of the stationary distribution of productivities z equals one. The computational algorithm is again unchanged except it imposes (60) and (61) rather than (11).

4 Mean Field Games (MFG)

4.1 Background

Mean field game theory is the study of strategic decision making in very large populations of small interacting agents. In HACT, all the agents are small but have impact on the asset prices when making saving decisions, so it is like a "game".

In continuous time a mean field game is typically composed by a Hamilton–Jacobi–Bellman equation that describes the optimal control problem of an individual and a Fokker–Planck equation that describes the dynamics of the aggregate distribution of agents. Under fairly general assumptions it can be proved that a class of mean field games is the limit as $N \to \infty$ of a N player Nash equilibrium.

4.2 Classical Example of MFG in Economics

We consider a large number of oil producers as price-takers in a perfectly competitive market. Each producer has a random initial reserve of oil $R_0 \sim m(0,\cdot)$. The distribution of oil reserve among producers is $m(t,\cdot)$. For each producer, the oil reserve $R(t)$ evolves with the production choice $q(t)$ together with a stochastic term as:

$$dR(t) = -q(t)dt + \nu R(t) dW_t \quad (*)$$

where ν measures the intensity of the noise. This intensity (or volatility) is supposed to be the same for all producers but the noise is proportional to the oil reserve. Then each producer makes the production choices $q(t)$ to optimize the profit:

$$\max_{(q(t))_t} \mathbb{E}\left[\int_0^\infty e^{-rt}(p(t)q(t) - C(q(t)))dt\right] \quad \text{s.t. } q(t) \geq 0, R(t) \geq 0$$

where:

- C is the production cost function: $C(q) = \alpha q + \beta \frac{q^2}{2}$
- the prices $p(t)$ are determined according to the demand-supply equilibrium

$$D(t,p) = We^{\rho t}p^{-\sigma} = \int m(t,R)q(t,R)dR$$

According to assumptions above, we introduce a Bellman function $u(t,R)$ as:

$$u(t,R) = \max_{(q(s))_{s\geq t}, q\geq 0} \mathbb{E}_t\left[\int_t^\infty (p(s)q(s) - C(q(s)))e^{-r(s-t)}ds\right]$$

$$dR(s) = -q(s)ds + \nu R(s)dW_s, \quad R(t) = R, \quad \forall s \geq t, R(s) \geq 0$$

We can discretize the above bellman equation to discrete from:

$$u(t,R) = \max_{(q(t)), q(t)\geq 0} [p(t)q - C(q(t))]\,\Delta t + (1 - r\Delta t)\mathbb{E}_t u(t+\Delta t, R_{t+\Delta t})$$

then we can derive:

$$0 = \Delta t\left[\max_{(q(s))_{s\geq t}, q\geq 0} [p(t)q - C(q(t))] + \frac{E_t u(t+\Delta t, R_{t+\Delta t}) - u(t,R)}{\Delta t} - rE(t+\Delta t, R_{t+\Delta t})\right]$$

and

$$0 = \max_{(q(s))_{s\geq t}, q\geq 0} [p(t)q - C(q(t))] + \frac{E_t u(t+\Delta t, R_{t+\Delta t}) - u(t,R)}{\Delta t} - rE(t+\Delta t, R_{t+\Delta t})$$

Using Taylor's expansion and Eq.(*), we obtain the following Hamilton-Jacobi-Bellman (HJB) equation:

$$\partial_t u(t,R) + \frac{\nu^2}{2}R^2\partial^2_{RR}u(t,R) - ru(t,R) + \max_{q\geq 0}\left(p(t)q - C(q) - q\partial_R u(t,R)\right) = 0$$

The Hamiltonian of this problem is $\max_{q\geq 0}\left(p(t)q - C(q) - q\partial_R u(t,R)\right)$. Using the quadratic cost function as above, the optimal control is given by:

$$q^*(t,R) = \left[\frac{p(t) - \alpha - \partial_R u(t,R)}{\beta}\right]_+$$

where $q^*(t, R)$ represents the instantaneous production at time t of a producer with an oil reserve R at time t. It's important to notice that R is the reserve at time t, not the initial reserve.

The Hamilton-Jacobi-Bellman equation can be rewritten with the optimal production:

$$\partial_t u(t, R) + \frac{\nu^2}{2} R^2 \partial_{RR}^2 u(t, R) - ru(t, R) + \frac{1}{2\beta} \left[(p(t) - \alpha - \partial_R u(t, R))_+ \right]^2 = 0$$

For the distribution of oil reserves at time t $m(t, R)$, it is initially given by $m_0(\cdot)$ and then transported by the optimal production decisions of the agents $q^*(t, R)$. The transport equation (Kolmogorov Forward or Fokker–Planck equation) is:

$$\partial_t m(t, R) + \partial_R \left(-q^*(t, R) m(t, R) \right) = \frac{\nu^2}{2} \partial_{RR}^2 \left[R^2 m(t, R) \right]$$

with $m(0, \cdot) = m_0(\cdot)$

These HJB and KF equations are the two classical coupled PDEs of the mean field game theory.

Reference

Achdou, Y., Han, Jiequn, Lasry, J.M., Lions, P.L. and Moll, B., 2017. Income and wealth distribution in macroeconomics: A continuous-time approach (No. w23732). National Bureau of Economic Research.

Achdou, Y., Han, Jiequn, Lasry, J.M., Lions, P.L. and Moll, B., 2017. Online Appendix: Numerical Methods for "Income and wealth distribution in macroeconomics: A continuous-time approach".`http://www.princeton.edu/~moll/HACTproject/HACT_Numerical_Appendix.pdf`

Guéant, O., Lasry, J.M. and Lions, P.L., 2011. Mean field games and applications. In Paris-Princeton lectures on mathematical finance 2010 (pp. 205-266). Springer, Berlin, Heidelberg.

Macroeconomic Analysis with Machine Learning and Big Data

Lecture 8: Heterogeneous Agent New Keynesian (HANK) Models

Instructor: Yucheng Yang *Scribe: Xiang Shi, Yuzhi Hao,Zhou Xing*

Disclaimer: *These notes have not been subjected to careful scrutiny and may contain errors. Please use it at your own risk.*

1 Introduction

Key questions always to keep in mind: Does heterogeneity matters? Why heterogeneity matters?

1.1 RANK (Representative Agent New Keynesian)

1.1.1 Model Setup

$$\max \int_0^\infty e^{-\rho t} U(C_t)\, dt, \quad U(C) = \frac{C^{1-\gamma}}{1-\gamma}, \quad \gamma > 0$$
$$s.t.\, \dot{a}_t = Y_t + r_t a_t - c_t$$

where Y_t is the output following the production function $Y_t = N_t$ and N_t is the labor supply. r_t is the interest rate and a_t is the asset holding. C_t is the consumption.

1.1.2 Monetary Transmission in RANK

We can solve this problem to get the Euler equation

$$\frac{\dot{C}_t}{C_t} = \frac{1}{\gamma}(r_t - \rho)$$

Solving this ODE we will get the path of consumption

$$C_t = \overline{C} \exp\left(-\frac{1}{\gamma}\int_t^\infty (r_s - \rho)\, ds\right)$$

To close the model we exogeneously assume a interest rule such that

$$r_t = \rho + e^{-\eta t}(r_0 - \rho), \quad t \geq 0$$

This formula collapses to a simple expression for the elasticity of initial consumption to the initial change in the interest rate

$$\frac{d \log C_0}{dr_0} = -\frac{1}{\gamma\eta}$$

where η is the speed of interesting to decentralized and γ is the elasticity of substitution.
Total differential of $log(C_0)$ is

$$d\log C_0 = -\frac{1}{\gamma}\int_0^\infty e^{-\rho t} dr_t dt - \frac{\rho}{\gamma}\int_0^\infty e^{-\rho t}\int_t^\infty dr_s ds dt$$

where the first term is the direct response to r and second term is the indirect effects due to Y
By giving the interest path we can solving this analytically

$$-\frac{d\log C_0}{dr_0} = \frac{1}{\gamma\eta}\left(\underbrace{\frac{\eta}{\rho+\eta}}_{\text{direct response to } r} + \underbrace{\frac{\rho}{\rho+\eta}}_{\text{indirect effects due to } Y}\right)$$

Normally the first term, direct response, is 99% and the second term, indirect effects, is 1% calculated by the model. However is result is not supportted by the data.

1.2 TANK (Two Agent New Keynesian Models)

1.2.1 Model setting

The setup is identical, except that we assume that a fraction Λ of households consume their entire current income. Therefore they will not directly response to the interest rate since no matter what they will always consume all of their income.

1.2.2 Monetary Transmission in TANK

By the same logic and method we can solve this problem to get

$$-\frac{d\log C_0}{dr_0} = \frac{1}{\gamma\eta}\left(\underbrace{(1-\Lambda)\frac{\eta}{\rho+\eta}}_{\text{direct response to } r} + \underbrace{\left((1-\Lambda)\frac{\rho}{\rho+\eta} + \Lambda\right)}_{\text{indirect due to } Y}\right)$$

A reasonable estimate for the proportion of handto- mouth households in the United States is 0.3 (Kaplan, Violante and Weidner, 2014). Thus, in TANK the share of direct effects is roughly 0.7. The overall effect in TANK is the same as in RANK because the addition of hand-to-mouth households decreases direct effects and increases indirect effects by the same magnitude. However, the results estimated by TANK still do not match the real data well.

2 HANK (Heterogeneous Agent New Keynesian Models)

2.1 Model Setup

The innovation of HANK model is that households face uninsurable idiosyncratic income risk which they can self-insure through two savings instrument with different degree of liquidity and return rates.

2.1.1 Household

$$\max_{\{c_t,l_t,k_t\}} E_0 \int_0^\infty e^{-(\rho+\zeta)t} u\left(c_t, \ell_t\right) dt$$

s.t.

$$\begin{aligned}
\dot{b}_t &= \underbrace{\left(1-\tau_t\right) w_t z_t \ell_t}_{\text{Income}} + \underbrace{r_t^b\left(b_t\right) b_t}_{\text{Capital}} + \underbrace{T_t}_{\text{Transfer}} - \underbrace{d_t}_{\text{Deposit}} - \underbrace{\chi\left(d_t, a_t\right)}_{\text{Transaction cost}} - \underbrace{c_t}_{\text{Consumption}} \\
\dot{a}_t &= r_t^a a_t + d_t \\
b_t &\geq -\underline{b}, \quad a_t \geq 0
\end{aligned}$$

where ζ is rate of an exodogenous death Poisson process,and upon death households give birth to an offspring with wealth a_0. z_t is idiosyncratic labor productivity following a log jump drift process to be specified later. b_t refers to household's liquid asset with interest rate r_t^b, and a_t refers to household's illiquid asset with interest rate r_t^a. Labor earnings are taxed at rate τ_t.

Transaction cost $\chi(d, a)$ is given by

$$\chi(d, a) = \chi_0|d| + \chi_1 \left|\frac{d}{a}\right|^{X_2} a$$

$$\chi_0 > 0, \chi_1 > 0, \chi_2 > 1$$

which ensures that $|d_t| < \infty$.

Households maximize life-long utility by choosing decision path $\{c_t, l_t, k_t\}_{t\geq 0}$, taking as given the time path of prices and policies $\{\Gamma_t\}_{t\geq 0} := \left\{r_t^b, r_t^a, w_t, \tau_t, T_t\right\}_{t\geq 0}$. Similar to HACT, we will derive the HJB and KF equations for the household problem.

$$V(a, b, y, t) = \max_{\{c_t,l_t,d_t\}} U(c, l)\Delta t + (1-(\rho+\zeta)\Delta t)(V(a_{t+\Delta t}, b_{t+\Delta t}, y_{t+\Delta t}, t+\Delta t)$$

where $y_{i,t} \equiv log z_{i,t}$ follows a "jumps-drift process":

$$dy_{i,t} = -\beta y_{i,t} dt + dJ_{i,t}.$$

Jumps arrive at a Poisson arrival rate λ. Conditional on a jump, a $y'_{it} \sim \mathcal{N}(0, \sigma^2)$ is drawn. Denote ϕ as density of a normal distribution with variance σ^2. Rearrange and divide both sides by Δt,

$$0 = \max_{c,\ell,d} U(c,l) + \frac{V_{t+\Delta t} - V_t}{\Delta t} - (\rho+\zeta)V_{t+\Delta t}$$

Take $\Delta t \to 0$:

$$0 = \max_{c,\ell,d} U(c,l) + \partial_a V \cdot \dot{a} + \partial_b V \cdot \dot{b} + \partial_y V \cdot \dot{y} + \partial_t V - (\rho+\zeta)V$$

HJB in the steady state equilibrium:

$$\begin{aligned}
(\rho+\zeta)V(a,b,y) = & \max_{c,\ell,d} u(c,\ell) + V_b(a,b,y)\left[(1-\tau)we^y\ell + r^b(b)b + T - d - \chi(d,a) - c\right] \\
& + V_a(a,b,y)\left(r^a a + d\right) \\
& + V_y(a,b,y)(-\beta y) + \lambda \int_{-\infty}^{\infty} (V(a,b,x) - V(a,b,y))\phi(x)dx
\end{aligned}$$

For the KF equation, define:

$$\begin{aligned}
G(a,b,y,t) &\equiv Pr(\tilde{a}_t \le a, \tilde{b}_t \le b, \tilde{y}_t \le y) \\
s^a &= ra + d \\
s^b &= \underbrace{(1-\tau)\omega Zl}_{\text{Income}} + \underbrace{r^b b}_{\text{Capital}} + \underbrace{T}_{\text{Transfer}} - \underbrace{d}_{\text{Deposit}} + \underbrace{\chi(d,a)}_{\text{Transaction cost}} - \underbrace{C}_{\text{Consumption}} \\
dy_{it} &= -\beta y_{it} dt + dJ_{it}
\end{aligned}$$

$$\begin{aligned}
G(a,b,y,t+\Delta t) &= Pr(\tilde{a}_{t+\Delta t} \le a, \tilde{b}_{t+\Delta t} \le b, \tilde{y}_{t+\Delta t} \le y) \\
&= \zeta\Delta t \Phi(a,b,y) + Pr(\tilde{a}_t \le a - \Delta t s^a, \tilde{b}_t \le b - \Delta t s^b, \tilde{y}_{t+\Delta t} \le y) \cdot (1-\zeta\Delta t) \\
&= \zeta\Delta t \Phi(a,b,y) + [Pr(\tilde{a}_t \le a - \Delta t s^a, \tilde{b}_t \le b - \Delta t s^b, \tilde{y}_t \le y + \Delta t \beta y_{it}) \cdot (1-\lambda\Delta t) \\
&\quad + Pr(\tilde{a}_t \le a - \Delta t s^a, \tilde{b}_t \le b - \Delta t s^b) \cdot \lambda\Delta t \cdot \int_{-\infty}^{Y}\int_{-\infty}^{+\infty} \varphi(y) g(\tilde{a}_t, \tilde{b}_t, x) dx dy](1-\zeta\Delta t) \\
&= G(a - \Delta t s^a, b - \Delta t s^b, y + \Delta t \beta y, t) \cdot (1-\lambda\Delta t) \cdot (1-\zeta\Delta t) + \zeta\Delta t \Phi(a,b,y) \\
&\quad + \lambda\Delta t \cdot \int_{-\infty}^{Y}\int_{-\infty}^{+\infty} \phi(y) G(a - \Delta t s^a, b - \Delta t s^b, x) dx dy
\end{aligned}$$

Subtract $G(a,b,y,t)$ from both sides, take the limit as $\Delta t \to 0$ and take derivative with respect to a, b, y, we obtain the KF equation in the steady state equilibrium:

KF in the steady state equilibrium:

$$\begin{aligned}
0 = & -\partial_a\left(s^a(a,b,y)g(a,b,y)\right) - \partial_b\left(s^b(a,b,y)g(a,b,y)\right) \\
& - \partial_y(-\beta y g(a,b,y)) - \lambda g(a,b,y) + \lambda\phi(y)\int_{-\infty}^{\infty} g(a,b,x)dx \\
& - \zeta g(a,b,y) + \zeta\delta\left(a - a_0\right)\delta\left(b - b_0\right) g^*(y)
\end{aligned}$$

where δ is the Dirac delta function, (a_0, b_0) are starting assets and $g^*(y)$ is the stationary distribution of y. The last term implies that so as to have the equilibrium to be achieved, after the origin household dies, the new-born households, with initial wealth, should have the stationary distribution of income.

2.1.2 Producers

Mathematically, the production side provides conditions to set the prices and close the model.

Final-Good Producers uses a continuum of intermediate inputs indexed by $j \in [0,1]$ and produce with constant elasticity of substitution ε (CES):

$$Y_t = \left(\int_0^1 y_{j,t}^{\frac{\varepsilon-1}{\varepsilon}} \, dj \right)^{\frac{\varepsilon}{\varepsilon-1}}$$

Cost minimization[1]gives the demand for intermediate good j :

$$y_{j,t}\left(p_{j,t}\right) = \left(\frac{p_{j,t}}{P_t} \right)^{-\varepsilon} Y_t, \quad \text{where} \quad P_t = \left(\int_0^1 p_{j,t}^{1-\varepsilon} dj \right)^{\frac{1}{1-\varepsilon}}$$

Intermediate Good Producers produce intermediate good j in monopolistically competitive market by:

$$y_{j,t} = k_{i,t}^{\alpha} n_{i,t}^{1-\alpha}$$

Again, cost minimization:

$$\min_{\{n_{j,t}, k_{j,t}\}} \omega_t n_{j,t} + r_t^k k_{j,t}$$

s.t.

$$y_{j,t} = k_{j,t}^{\alpha} n_{j,t}^{1-\alpha}$$
$$y_{j,t} = \left(\frac{p_{j,t}}{P_t} \right)^{-\varepsilon} Y_t$$

from which we can solve for $k_{j,t}, n_{j,t}$ and this also implies that marginal cost m_t is given by:

$$m_t = \left(\frac{r_t^k}{\alpha} \right)^{\alpha} \left(\frac{w_t}{1-\alpha} \right)^{1-\alpha}$$

[1]Construct $\mathcal{L} = \min \sum_{j=0}^{1} P_j y_j - \lambda[\left(\int_0^1 y_{j,t}^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}} - Y_t]$, let $\frac{\partial \mathcal{L}}{\partial y_{j,t}} = 0$ and get $p_j = \lambda Y^{\frac{1}{\varepsilon}} \cdot y_j^{-\frac{1}{\varepsilon}}$. Take two indexes of products j_1 and j_2 and get $y_{j_2} = \left(\frac{p_{j_1}}{p_{j_2}} \right)^{\varepsilon} y_{j_1}$. Taking the $\frac{\varepsilon-1}{\varepsilon}$ power of both sides and integral by dj_2 yields $\left(\int_0^1 y_{j_2}^{\frac{\varepsilon-1}{\varepsilon}} d_{j_2} \right)^{\frac{\varepsilon}{\varepsilon-1}} = \left(y_{j_1}^{\frac{\varepsilon-1}{\varepsilon}} p_{j_1}^{\varepsilon-1} \right)^{\frac{\varepsilon}{\varepsilon-1}} \cdot \left(\int_0^1 p_{j_2}^{1-\varepsilon} d_{j_2} \right)^{\frac{\varepsilon}{\varepsilon-1}}$.

We define cost adjustment cost:

$$\Theta_t\left(\frac{\dot{p}_t}{p_t}\right) = \frac{\theta}{2}\left(\frac{\dot{p}_t}{p_t}\right)^2 Y_t = \frac{\theta}{2}\pi_t^2 Y_t$$

Since producers are in monopolistically competitive market, the profit for representative intermediate good producers can be defined as:

$$\Pi_t(p_t) \equiv (\frac{p_t}{P_t} - m_t)(\frac{p_t}{P_t})^{-\varepsilon} Y_t - \frac{\theta}{2}\pi_t^2 Y_t$$

Households' illiquid assets can be invested into two assets: capital k_t and equity share of intermediate firms s_t, which implies $a_t = k_t + q_t s_t$. The dynamics of capital and equity satisfy:

$$\dot{k}_t + q_t \dot{s}_t = (r_t^k - \delta)k_t + \Pi_t s_t + d_t$$

Since households can switch costlessly between capital and equity, the return of both assets should be equal, which implies

$$\frac{\Pi_t + \dot{q}_t}{q_t} = r_t^k - \delta = r_t^a$$

And **this connects** r_t^k **with** r_t^a.

2.1.3 Monetary Authority

According to the Taylor Rule:

$$i_t = \overline{r}^b + \phi\pi_t + \epsilon_t$$

where ϵ_t is the monetary shock. And according to the Fisher Equation:

$$r_t^b = i_t - \pi_t$$

And **this connects** r_t^b **with** ϵ_t.

2.1.4 Government

The governmental budget balance condition gives us:

$$\dot{B}_t^g + G_t + T_t = \tau_t \int w_t z \ell_t(a,b,z) d\mu_t + r_t^b B_t^g$$

where μ_t is the measure of households over (a, b, z). To note, outside of steady state, the fiscal instrument to balance the budget can be either T_t, τ_t or G_t, and the responses of other variables would depend on which fiscal instrument to use.

2.2 Equilibrium

- Decisions: $\{a_t, b_t, c_t, d_t, \ell_t, n_t, k_t\}_{t\geq 0}$
- States: $\{\Gamma_t\}_{t\geq 0} := \left\{\omega_t, r_t^k, r_t^a, r_t^b, q_t, \pi_t, \tau_t, \boldsymbol{T}_t, \boldsymbol{G}_t, \boldsymbol{B}_t, \mu_t\right\}_{t\geq 0}$
  - Liquid asset clears: $B_t^h + B_t^g = 0$
  - Illiquid asset clears: $K_t + q_t = A_t$, with the total number of shares normalized to 1
  - good market clears: $Y_t = C_t + I_t + G_t + \Theta_t + \chi_t + \kappa \int \max\{-b, 0\} d\mu_t$, the last term reflects borrowing costs.

3 Monetary Transmission in HANK

Consider an expansionary monetary shock ($\epsilon_0 < 0$). Recall from Section 2.1.3, ϵ_t is directly connected with liquid assets' interest rate. The monetary shock also affect other state variables in Γ_t.

We define aggregate consumption as

$$C_t\left(\{\Gamma_t\}_{t\geq 0}\right) = \int c_t\left(a, b, z; \{\Gamma_t\}_{t\geq 0}\right) d\mu_t$$

which can be total differentiated into:

$$dC_0 = \underbrace{\int_0^\infty \frac{\partial C_0}{\partial r_t^b} dr_t^b dt}_{\text{direct effect } (\approx 19\%)} + \underbrace{\int_0^\infty \left(\frac{\partial C_0}{\partial w_t} dw_t + \frac{\partial C_0}{\partial r_t^a} dr_t^a + \frac{\partial C_0}{\partial \tau_t} d\tau_t + \frac{\partial C_0}{\partial T_t} dT_t\right) dt}_{\text{indirect effects} (\approx 81\%)}$$

4 Solving the steady state equilibrium

- Step 1: Make initial guess for $\{K, N, r^b\}$ (or guess prices $\{r^a, r^b, \omega\}$ and solve the remaining in Step 2).
- Step 2: Using the relations we get from the producer's problem to solve for $\{\omega, r^a, \Pi, \tau/T/G\}$ as necessary input to solve the HJB.
- Step 3: Given $\omega, r^a, r^b, \Pi, \tau/T/G$, solve the HJB equation to get policy function a, b, c, l, d with the HACT method.
- Step 4: Given policy functions, solve the KF equation for distribution g with the HACT method.
- Step 5: Check if market clear conditions are satisfied. If not, update the values in the Step 1 according to the difference.

5 Summary: Comparison between RANK and HANK

- RANK views:
  - High sensitivity of C to r: intertemporal substitution
  - Low sensitivity of C to Y: the representative agent is permanent income hypothesis (PIH) consumer
- HANK views:
  - Low sensitivity to r: income effect of wealthy offsets substitution effect
  - High sensitivity to Y: sizable share of hand-to-mouth agents

6 The Road Ahead

In both HACT and HANK models, the labor productivity z is idiosyncratic for households, while the aggregation remains unchanged over time. What if the aggregated Y_t and Z_t are stochastic? The problem becomes much more complicated mathematically. Economists have proposed methods like perturbation to solve such problems, and other tools like deep neural networks might be helpful to further explore the non-linear dynamics of the distribution. We will discuss some recent work in the next class.

Reference

Kaplan, G., Moll, B. and Violante, G.L., 2018. Monetary policy according to HANK. American Economic Review, 108(3), pp.697-743.

Kaplan, G., Moll, B. and Violante, G.L., 2018. Equilibrium Conditions and Algorithm for Numerical Solution of Kaplan, Moll and Violante (2017) HANK Model. `https://benjaminmoll.com/wp-content/uploads/2019/07/HANK_algorithm.pdf`

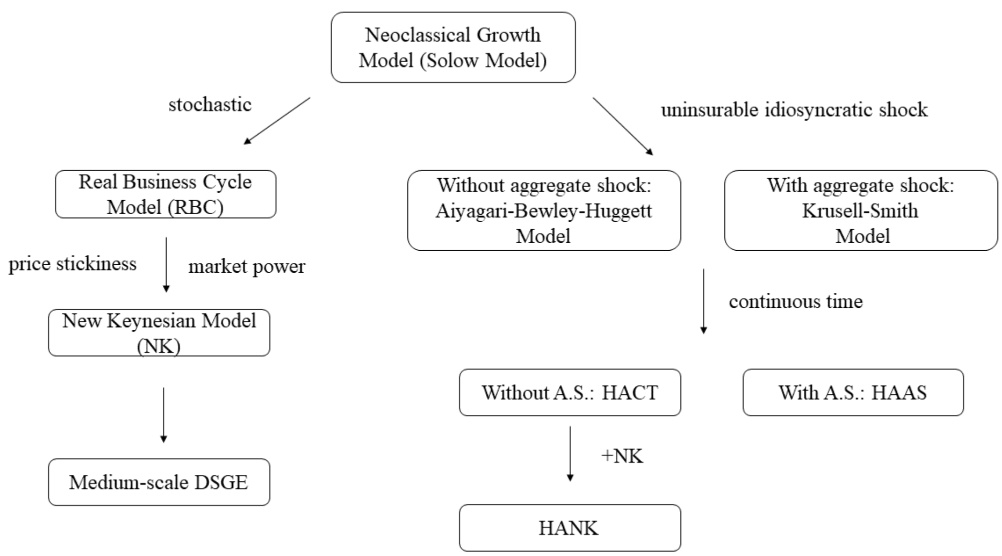
Macroeconomic Analysis with Machine Learning and Big Data

Lecture 9: HACT with Aggregate Shocks

Instructor: Yucheng Yang *Scribe: Jieyu Guo, Zehou Zhang*

Disclaimer: *These notes have not been subjected to careful scrutiny and may contain errors. Please use it at your own risk.*

1 Recap: Structural Models in Macroeconomics



2 Model Setup

I follow Ahn *et al.* (2018, NBER Macro Annual) to introduce the basic model setup of HACT with aggregate shocks. I will introduce their solution method, as well as the solution method proposed by Fernandez-Villaverde *et al.* (2019) in this note.

- Household (HH): heterogenous over asset a and effective productivity z

$$E_0 \int_0^\infty e^{-\rho t} \frac{C_{jt}^{r\theta}}{1-\theta} dt$$

$$a_{jt} = w_t z_{jt} + r_t a_{jt} - C_{jt}$$

$$z_{jt} \in \{z_L, z_H\}, z_L < z_H$$

$$\text{Assume } \overline{N} \equiv \int_0^1 z_{jt} d_j \text{ is constant } \forall t$$

- Firm: representative in competitive market.

$$Y_t = e^{H_t} K_t^{\alpha} N_t^{1-\alpha}$$

  Log productivity H_t is the only source of **aggregate shocks** in this model and follow the Ornstein-Uhlenbeek process:

$$dH_t = -\eta H_t dt + \sigma dW_t$$

  As firms are in the competitive market, the factor prices are:

$$r_t = \alpha e^{H_t} K_t^{\alpha-1} \overline{N}_t^{1-\alpha}$$

$$w_t = (1-\alpha) e^{H_t} K_t^{\alpha} \overline{N}^{-\alpha}$$

- Hamilton-Jacobi-Bellman (HJB) equation:
  - individual state variables: a, z_j
  - aggregate state variables: $H, g(a, z)$
  - Remark: in HACT, $g_t(a, z)$ is deterministic given initial conditions, so $g_t(a, z)$ was not in the state variables. With aggregate shocks, $g_t(a, z)$ is stochastic and have many possible realizations, so we have it in the state variable.
  - The fully recursive formulation of the HH problem (**the HJB equation**) is:

$$\begin{aligned} \rho v(a, z, g_t, H_t) = & \max_c u(c) + \partial_a v(a, z, g_t, H_t)[\omega(g, H)Z + r(g_t, H_t)a - c] \\ & + \lambda_z [v(a, z', g_t, H_t) - v(a, z, g_t, H_t)] \\ & + \partial_H v(a, z, g_t, H_t)(-\eta H_t) + \frac{1}{2}\partial^2_{HH} v(a, z, g_t, H_t)\sigma^2 \\ & + \int \frac{\delta v(a, z, g_t, H_t)}{\delta g(a, z)} (\kappa_H g)(a, z) da dz \end{aligned} \tag{1}$$

    where $\frac{\delta v(a,z,g,H)}{\delta g(a,z)}$ is the functional derivative, and κ_H is the Kolmogorov Forward operator to be defined later.

- **Kolmogorov Forward (KF) Equation**:

$$\kappa_H g_t(a, z) = \frac{dg_t}{dt}$$

  where the Kolmogorov Forward operator κ_H is defined as

$$(\kappa_H g)(a, z) = -\partial_a [s(a, z, g_t, H_t) g(a, z)] - \lambda_z g(a, z') + \lambda_{z'} g(a, z')$$

Comparing to HACT, the HJB equation here with the functional derivative component is too complicated to solve. Fortunately, we can simply this formulation with expectation terms getting in. Using Ito's formula, we have

$$dv(a,z,H_t,g_t) = \partial_H v(a,z,g_t,H_t)(-\eta H_t)dt + \frac{1}{2}\partial_{HH}v(a,z,g_t,H_t)\sigma^2 dt + \sigma\partial_H v(a,z,g_t,H_t)dW_t + \int \frac{\delta v}{\delta g}\kappa_H g_t(a,z)dadzdt \tag{2}$$

Taking expectation to get rid of the Brownian motion, we have

$$\mathbb{E}_t dv(a,z,g_t,H_t) = \partial_H v(a,z,g_t,H_t)(-\eta H_t)dt + \frac{1}{2}\partial_{HH}v(a,z,g_t,H_t)\sigma^2 dt + \int \frac{\delta v}{\delta g}\kappa_H g_t(a,z)dadzdt \tag{3}$$

Plug into Equation 1, **the HJB equation** can be finally written as:

$$\rho v_t(a,z) = \max_c u(c) + \partial_a v_t(a,z)(\omega_t Z + r_t a - c) + \lambda_z(v_t(a,z') - v_t(a,z)) + \frac{1}{dt}\mathbb{E}_t dv_t(a,z) \tag{4}$$

This formulation with expectation term reminds us of the RBC models. It turns out that one of the solution method approximate the problem with a linear rational expectation system problem and solve it with the techniques we have used before.

3 Solution Method I: Linearization around the Steady State Equilibrium

The solution method proposed by Ahn *et al.* (2018) consists of three steps.

- Step 1: Solve the Steady State Equilibrium without Aggregate Shock.

  This step solve the steady state equilibrium without aggregate shocks (i.e. $H_t \equiv 0$) but with idiosyncratic shocks. The HJB and KF equations, and the market clear condition are

$$\rho v(a,z) = \max_c u(c) + \partial_a v(a,z)(\omega Z + ra - c) + \lambda_z(v(a,z') - v(a,z))$$

$$0 = -\partial_a[s(a,z)g(a,z)] - \lambda_z g(a,z') + \lambda_{z'} g(a,z')$$

$$K = \int a g(a,z)dadz$$

  It can be solved with the same method as HACT.

- Step 2: Linearize Equilibrium Conditions:
  This step is to compute a first-order Taylor expansion of the model's discretized equilibrium conditions around steady state.

- discretized equilibrium for finite difference:

$$\begin{aligned}
\rho\overrightarrow{v_t} &= \overrightarrow{u}(\overrightarrow{v_t}) + A(\overrightarrow{v_t}, \overrightarrow{p_t})\overrightarrow{v_t} + \frac{1}{d_t}E_t dv_t \\
\frac{d\overrightarrow{g_t}}{d_t} &= \overrightarrow{A}(\overrightarrow{v_t}; \overrightarrow{p_t})^T\overrightarrow{g_t} \\
dH_t &= -\eta H_t dt + \sigma dW_t \\
\overrightarrow{p_t} &= \overrightarrow{F}(\overrightarrow{g_t}, \overrightarrow{H_t})
\end{aligned} \tag{5}$$

- first-order Taylor expansion

$$E_t \begin{bmatrix} d\hat{v}_t \\ d\hat{g}_t \\ dZ_t \\ 0 \end{bmatrix} = \begin{bmatrix} B_{vv} & 0 & 0 & B_{vp} \\ B_{gv} & B_{gg} & 0 & B_{gp} \\ 0 & 0 & -\eta & 0 \\ 0 & B_{pg} & B_{pH} & -I \end{bmatrix} \times \begin{bmatrix} \hat{v}_t \\ \hat{g}_t \\ H_t \\ \hat{p}_t \end{bmatrix} d_t \tag{6}$$

- plug the pricing equation $\hat{p}_t = B_{pg}\hat{g}_t + B_{pZ}Z_t$ into the remaining equations of the system:

$$E_t \begin{bmatrix} d\hat{v}_t \\ d\hat{g}_t \\ dZ_t \end{bmatrix} = \underbrace{\begin{bmatrix} B_{vv} & B_{vp}B_{pg} & B_{vp}B_{pH} \\ B_{gv} & B_{gg} + B_{gp}B_{pg} & B_{gp}B_{pH} \\ 0 & 0 & -\eta \end{bmatrix}}_{\mathbf{B}} \times \begin{bmatrix} \hat{v}_t \\ \hat{g}_t \\ H_t \end{bmatrix} d_t \tag{7}$$

- Step 3: Solve Linear System
  - Schur decomposition of the matrix B to identify the stable and unstable roots of the system.
  - If the Blanchard and Kahn (1980) condition holds, i.e. the number of stable roots equals the number of state variables $\hat{g}_t$ and Z_t, then we can compute the solution:

$$\begin{aligned}
\hat{v}_t &= D_{vg}\hat{g}_t + D_{vH}H_t, \\
\frac{d\hat{g}_t}{dt} &= (B_{gg} + B_{gp}B_{pg} + B_{gv}D_{vg})\hat{g}_t + (B_{gp}B_{pH} + B_{gv}B_{vH})H_t, \\
dH_t &= -\eta H_t dt + \sigma dW_t, \\
\hat{p}_t &= B_{pg}\hat{g}_t + B_{pH}H_t.
\end{aligned} \tag{8}$$

- Approximated Solution

$$v_t(a_j, z_j) = v(a_j, z_j) + \sum_{k=1}^{I}\sum_{l=1}^{2} D_{vg}[i, j; k, l](g_t(a_k, z_l) - g(a_k, z_l)) + D_{vz}[i, j]H_t \tag{9}$$

Optimal consumption is then given by:

$$c_t(a_i, z_j) = (\partial_a v_t(a_i, z_j))^{-1/\theta}$$

4 Solution Method II: Deep Neural Network Approximation

Fernandez-Villaverde *et al.* (2019) proposes a different method that can better address the nonlinearities in the model. To deal with the functional derivative component in the original HJB equation as below,

$$\begin{aligned}\rho v(a,z,g_t,H_t) = \max_c u(c) &+ \partial_a v(a,z,g_t,H_t)[\omega(g,H)Z + r(g_t,H_t)a - c] \\ &+ \lambda_z[v(a,z',g_t,H_t) - v(a,z,g_t,H_t)] \\ &+ \partial_H v(a,z,g_t,H_t)(-\eta H_t) + \frac{1}{2}\partial^2_{HH} v(a,z,g_t,H_t)\sigma^2 \\ &+ \int \frac{\delta v(a,z,g_t,H_t)}{\delta g(a,z)}(\kappa_H g)(a,z)dadz\end{aligned}$$

they use finite moments $\{m_{1,H},...,m_{N,H};m_{1,L},...,m_{N,L}\}$ to approximate the whole distribution of $g(\cdot,\cdot)$, and use a deep neural network to approximate the nonlinear law of motion $h(\cdot)$ of those moments over time.

In the algorithm, every time we have a guess for the law of motion of those moments $h_n(\cdot)$. Then we solve the HJB given $h_n(\cdot)$ and simulate time series data for the moments $\{m_{1,H},...,m_{N,H};m_{1,L},...,m_{N,L}\}$. With the simulated time series, we learn a new law of motion $h_{n+1}(\cdot)$ until it converges in some sense.

5 Course Summary