Discussion of “Latent Heterogeneity in the Marginal Propensity to Consume” by Lewis, Melcangi, Pilossoph

Sara Casella, University of Pennsylvania

November 5, 2019
Modelling with Big Data and Machine Learning: Interpretability and Model Uncertainty

What is this paper about?

What is this paper about?

What is this paper about?

- Distribution of Marginal Propensity to Consume (MPC) very important object in macroeconomics
- But still a lot of uncertainty around it: what is the average MPC in the data? who has high MPC?
- So far, either ignore heterogeneity or try to address problem with quantile regressions
- This paper's innovation: repeat a well known study but using **fuzzy C-means (FMC)** algorithm to group individuals based on latent characteristics and develop its properties

Methodology and Contributions

- Contribute to both clustering and macro literature

- Extend use of Fuzzy C-means to a fully general regression model
- Show that FCM fits into a GMM framework (very well know in Econometrics)
  - Work out properties and prove consistency
  - Propose a fast computational method

- Give us the distribution of MPCs and unbiased estimate of average MPC (higher than previously thought)
- Idea of which dimensions of heterogeneity are important and how much we don't know
  - Best battery of controls explains only 13% of the variation in MPCs
  - Most correlated covariates are income and mortgage payment

When is it appropriate to use FCM?

- **Causal Random Forests** (Wager and Athey 2018, Chernozhukov et al. 2017)
  - Find out whether treatment effect is heterogeneous for different groups
  - Groups based on many observable covariates (-)
- **K-means clustering** (Bonhomme and Manresa 2015)
  - Groups individuals based on unobservable characteristics (+)
  - Does not perform well in short panels and not separated data (-)
- **Classifier LASSO** (Su et al. 2016)
  - Finds heterogeneous slopes in panel data and classifies individuals in latent groups (+)
  - Performs possibly well in short panels (+)
  - No mixed membership (-)

- Conclusion: groups based on unobservable characteristics + short panel + noisy not separated data = perfect application for FCM!

Size Sensitivity and Heteroskedasticity

$$J_m^{reg}(\Pi, \mu^{reg}, \theta) = \int\int \sum_{g=1}^{G} \left(\mu_g^{reg}(y|x;\theta)\right)^m \|y - \theta_g x\|^2 \, \Pi_{y|x}(dy|x)\Pi(dx)$$

- C-means clustering tends to create balanced clusters in terms of size and variance (Bensaid et al 1996, Bezdek 1973)
- With heteroskedasticity there might be some mis-classification
- Is this important? Check by simulation
- Solution: adjust by measures of integrity (Lin et al. 2014), possibilistic fuzzy c-means (Pal, Pal, Keller, Bezdek 2005)

Simple Example

Drivers of MPC Heterogeneity

- Find that total income and mortgage interest payment are positively correlated with MPC
- What about future expenses? Are the clusters similar?
- Breakdown by goods, entertainment/luxury vs non luxury
- Is the correlation with income due to debt repayment? (Shapiro and Slemrod 2009) Could link CCP/Equifax to CEX (zipcode level)

Conclusion

- Great paper with range of contributions
- Among strengths of the paper showing FCM fits into a GMM framework
- Food for thought for macroeconomists interested in heterogeneity
- $\Rightarrow$ Lots of possible application

References

Bensaid, A.M., L.O. Hall, J.C. Bezdek, L.P. Clarke. 1996. “Partially supervised clustering for image segmentation”. Pattern recognition.

Bezdek , J.C. 1973. “Cluster validity with fuzzy sets”. Taylor & Francis

Bonhomme, S. and E. Manresa 2015 “Grouped Patterns of Heterogeneity in Panel Data”. Econometrica, 83:1147–1184.

Chernozhukov V., M. Demirer, E. Duflo, and I. Fernandez-Val. 2017. “Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments”. arXiv e-prints, arXiv:1712.04802.

Lin, P, P. Huang, H. Kuo, Y.H. Lai. 2014. “A size-insensitive integrity-based fuzzy c-means method for data clustering Pattern Recognition” 47-5:2042-2056

References

Pal, N.R., K. Pal, J.M. Keller, J.C. Bezdek. 2005. “A possibilistic fuzzy c-means clustering algorithm Publisher” IEEE Transactions on Fuzzy Systems 13(4):517-530

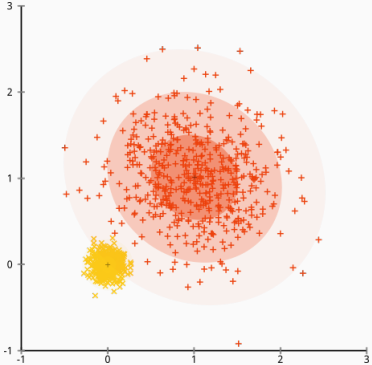
Shapiro, M.D. and J. Slemrod. 2009. “Did the 2008 Tax Rebates Stimulate Spending?” American Economic Review, 99, 374–79.

Lianjun, s, Z. Shi, P. Phillips. 2016. “Identifying Latent Structures in Panel Data”. Econometrica, Vol. 84-6:2215–2264

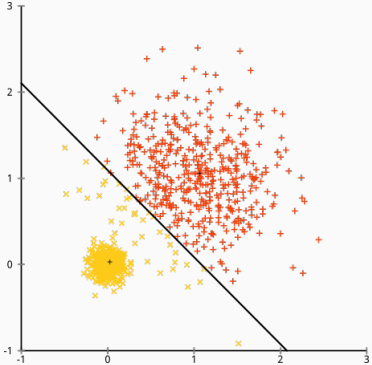
Wager, S and S. Athey. 2018. “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests” Journal of the American Statistical Society

Simple Example of Mis-classification

Back



(a) Gaussian Mixtures



(b) K-means clustering