Comments on
Estimating Policy Functions in Payments Systems Using Reinforcement Learning
Authors: Pablo S. Castro, Ajit Desai, Han Du, Rodney Garratt and Francisco Rivadeneyra

Discussant: Dr. Elizabeth Téllez
Senior Economist at CEMLA

Disclaimer: The opinions expressed here do not represent the views of CEMLA.
June 2021 – Mexico City
Summary of the paper

- The paper is very interesting with innovative methodology. Reinforcement Learning (RL) could promote the proper functioning of the Payment Systems (PS).

- Large-value payment systems (LVPS) generally settle in real-time. Banks choose the amount of liquidity provided to the payment systems
  - Given the costs of liquidity, banks and regulators fine tune their policy functions.

- Authors suggest a model with two agents who optimize liquidity using RL and learn an optimal policy
  - That minimizes the cost of processing their individual payments.
  - Yet, a more realistic LVPS would involve multiple participants and periods.

- The estimation of a policy function for LVPS using RL is then motivated by:
  1) Assisting policy-makers and payment system participants to define optimal initial liquidity at the lowest cost.
  2) Designing new payment systems.
Comments

- What other Machine Learning (ML) techniques did you consider to estimate this issue?
- For training you used the REINFORCE algorithm (policy gradient technique). There are a variety of methods to optimize an agent’s policy, even within RL, including:
  - Deterministic Policy Gradients
  - Evolutionary
  - Policy search
  - Model based
  - Imitation learning
- Why did you choose RL-REINFORCE?

Have you thought for your future research work to estimate more than two agents and periods with other ML techniques?
Comments

■ Overall comments
- The **applied game theory** literature is one of their contributions, as it is the starting point for the **RL exercise**. Prior work is the theoretical model of **Bech and Garratt (2003)**. 

- The **authors acknowledge** that:
  - Their model of the environment abstracts from **two important dimensions of PS**:
    - The **indivisibility of payments** and
    - The **interbank liquidity market**.

■ On the methodology
- The **RL agent learns** about their own environment. RL guides the participant behavior.
- Limitations of the **RL** models are the **non-analytic solutions**, but a **simulation-based optimization** is provided.
- It is important to have estimates of the sensitivity of the agent's best responses, at different levels of delay cost.
  - Because delay cost is unobservable to researchers and policy-makers.
  - How cost could be misleading without an interbank market?
Future work

We are curious to see your future work on:

- How would agents behave if they were knowing the initial liquidity and the inter-day payment, at the same time?

- Introduction of some realistic features of the payment system
  - Modeling more than two agents, a more complex scenario.
  - Non-divisible payments.
  - Considering (cheaper) interbank market liquidity.

Thinking of hybrid LVPS, how useful is this approach to design liquidity policy functions for (retail) fast payment systems? This could be a research extension.

Great and exciting fields to be explored!

Thank you!

itellez@cemla.org