An Efficient Liquidity Savings Mechanism

Darrell Duffie
Stanford GSB

Srisht Fateh Singh
University of Toronto

Chaojun Wang
Wharton School

The Future of Market Liquidity and Functioning Workshop
Federal Reserve Bank of New York
July 2026

Liabilities of the Federal Reserve System

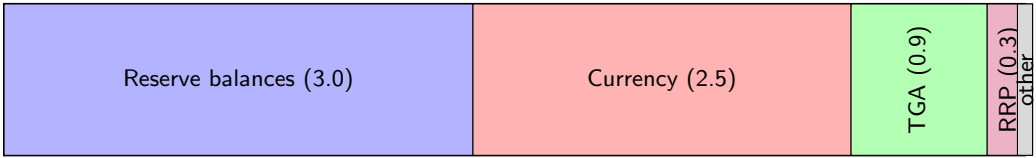
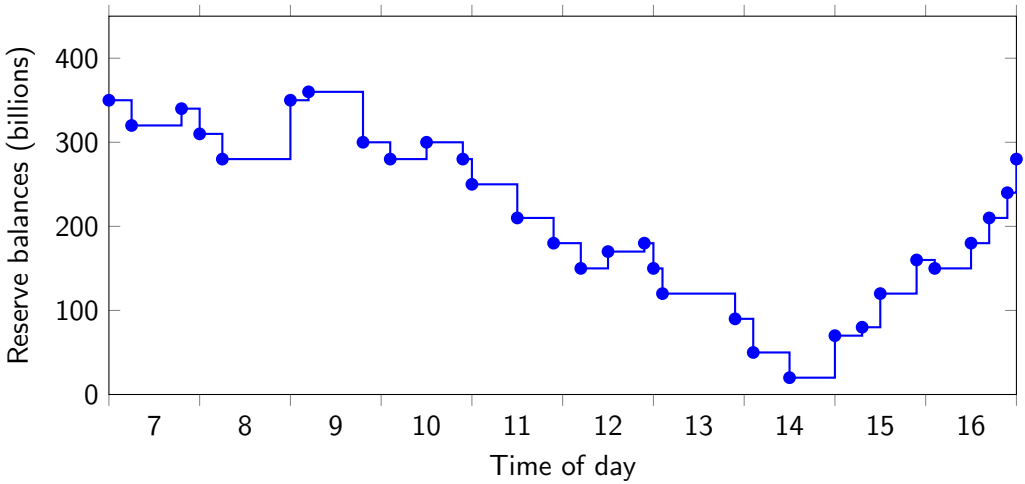
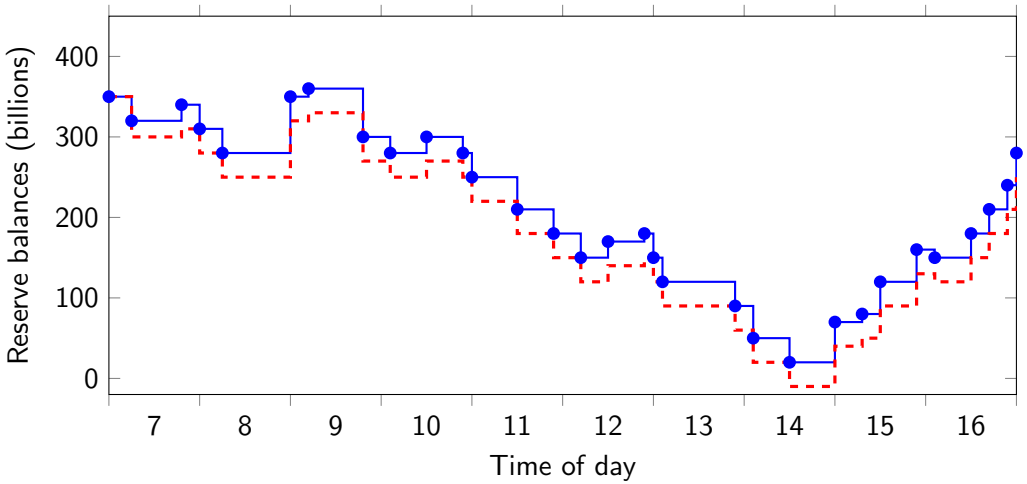


Figure: Liabilities of the Federal Reserve System in trillions of dollars, to scale, as of July 1, 2026. TGA is the Treasury General Account. RRP refers to the sum of the Foreign Official Reverse Repurchase pool ($336 billion) and the Overnight Reverse Repurchase Facility ($8 billion). Data: Federal Reserve Board, H.4.1.

A bank's possible intra-day path of reserve balances



Impact of providing 20 billion more in repo funding early in the day



Delayed payments to dealer banks impact rates

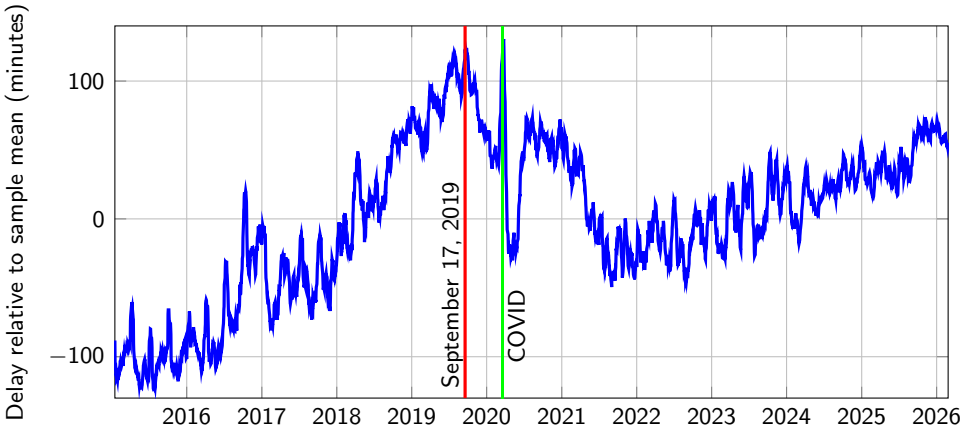


Figure: The 10-day trailing average of the time of day by which the ten largest repo-active bank holding companies had received half of their large ($\geq$ \$100 million) incoming payments, relative to sample mean, in minutes. Source: Copeland, Duffie, Wang, and Yang (2025). Data: Fedwire Funds Service.

Abundant reserves changed payment funding

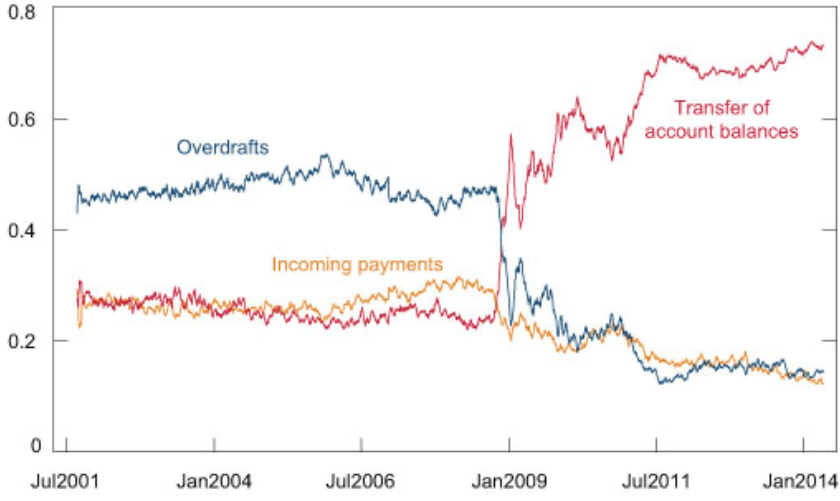


Figure: Fractional sources for aggregate Fedwire daily payments. Source: Garratt, Martin, McAndrews (2014), Federal Reserve Bank of New York.

The largest banks have almost stopped daylight overdrafts

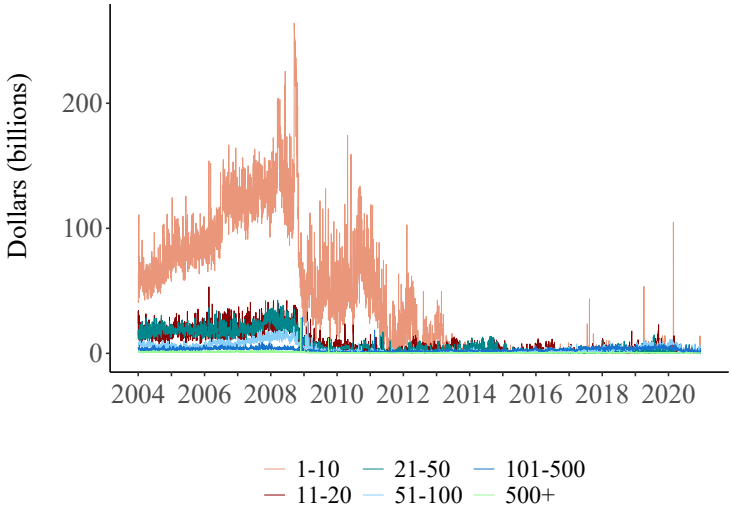


Figure: Overdrafts, banks ranked size. Source: Badev et al (2021), Federal Reserve Board.

An LSM can conserve reserves with multilateral offsets

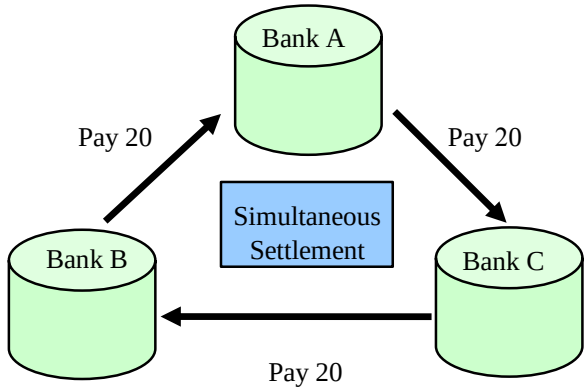


Figure: Multilaterally offsetting payments in the Bank of Japan's LSM, BOJNet. Figure source: Bank of Japan.

The model setting

- Banks $1, \ldots, n$ have initial balances $X_1, \ldots, X_n$ respectively.
- Bank i will pay $Q_{ij} \geq 0$ to bank j, of which p_{ij} is paid early. The rest is paid late.
- The information $(X_i, (Q_{ij})_{j \neq i})$ is private to bank i.
- Banks have a proportional cost ψ for overdrafts and $c < \psi$ for late payments.

Efficient liquidity savings mechanisms

Our LSM designs induce maximal participation by banks, and thus maximize total early payments and minimize total costs, in each of the following cases:

1. The LSM is constrained to use no reserves.
2. The LSM can use reserves but cannot cause overdrafts.
3. The LSM can cause overdrafts.

A maximally efficient LSM, subject to no use of reserve balances

- Bank i submits to the LSM a cap $\bar{p}_{ij}$ on its payment to bank j.
- The LSM makes payments that solve the linear program

$$\begin{aligned} \max_{\{p_{ij}\}} \quad & \sum_{i,j} p_{ij} \\ \text{s.t.} \quad & \sum_{j} p_{ij} - \sum_{j} p_{ji} \leq 0, \quad \textit{for all } i, \\ & 0 \leq p_{ij} \leq \bar{p}_{ij} \quad \textit{for all } (i,j). \end{aligned} \tag{1}$$

Proposition

For each bank i, maximal submission $(\bar{p}_{ij} = Q_{ij})_{j \neq i}$ to the LSM (1) is a weakly dominant strategy.

Theoretical impact of this LSM on the promptness of payments

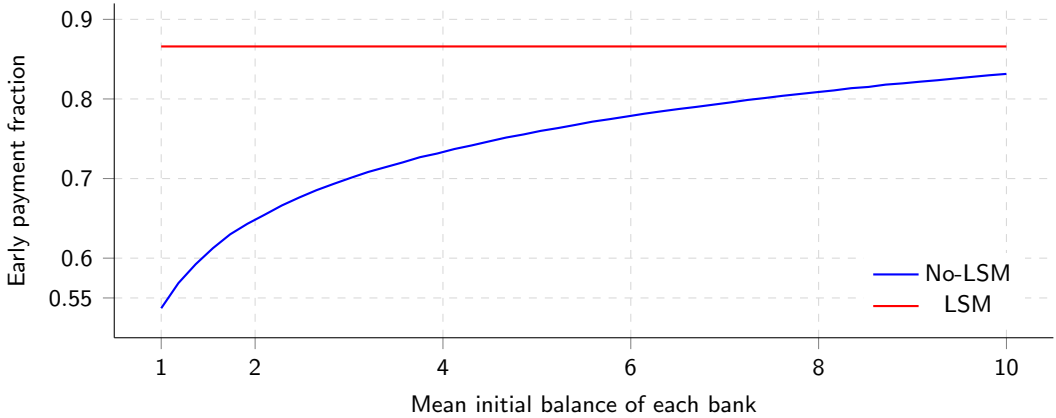


Figure: The perceived proportional cost of overdrafts is assumed to be twice that for late payments. See the paper for other parametric assumptions.

A maximally efficient LSM, subject to no overdrafts

- Bank i submits a cap $\bar{p}_{ij}$ on its LSM payment to bank j and a cap $\bar{p}_i^{\text{net}} \leq X_i$ on the LSM's net use of its reserves.
- Bank i pays or receives a proportional fee $\beta \leq c$ on net reserves received or paid.
- The LSM makes payments that solve the linear program:

$$\begin{aligned} \max_{\{p_{ij}:(i,j)\in G\}} \quad & \sum_{i,j} p_{ij} \\ \text{s.t.} \quad & \sum_j p_{ij} - \sum_j p_{ji} \leq \bar{p}_i^{\text{net}}, \qquad \textit{for all } i. \\ & 0 \leq p_{ij} \leq \bar{p}_{ij}, \qquad \textit{for all } (i,j). \end{aligned} \tag{2}$$

- The LSM uses a tie-breaking rule among LP solutions, explained in the paper.

LSM

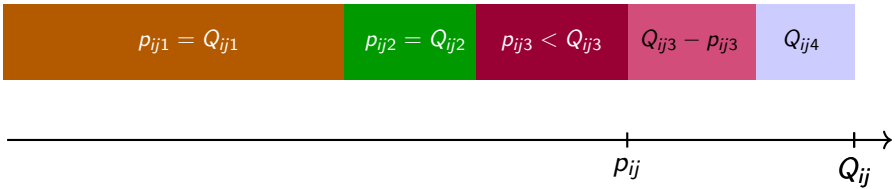
Proposition

The LP-based LSM, with fees and subsidies for net use of reserves, is budget balanced. Under this LSM, each bank i has a weakly dominant maximal submission, given by

$$\bar{p}_{ij}^{*} = Q_{ij}, \qquad \bar{p}_{i}^{*net} = \min\left(X_i, \sum_j Q_{ij}\right).$$

Under maximal submissions, the induced LSM payment outcome is efficient, subject to no overdrafts.

Alert: One of the payments from bank i to j bank is partially delayed



LSM performance depends on the core-periphery structure of banking

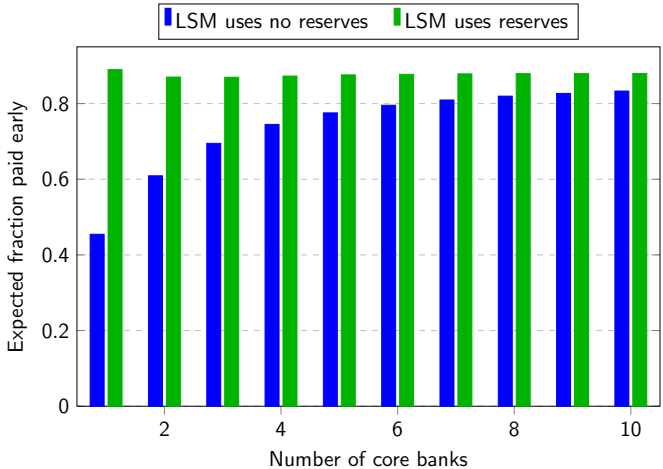


Figure: 20 banks, with total reserves of core banks held fixed. See the paper for the probability distribution of payment sizes.

Early payment fraction depends on the supply of reserves

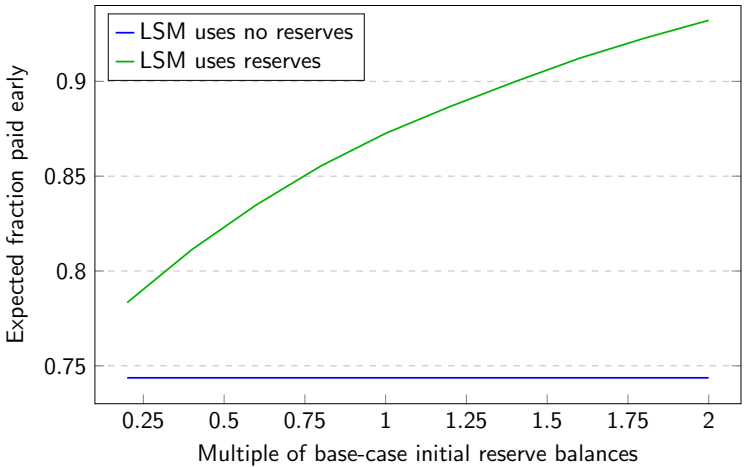


Figure: 20 banks, with 4 core banks.

Appendix

Liquidity savings mechanisms reduce the need for reserves

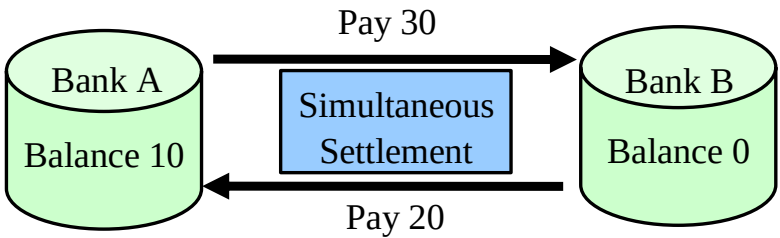


Figure: Bilaterally offsetting payments in the Bank of Japan's LSM, BOJNet. Figure source: Bank of Japan.

The LSM used by Bank of Japan

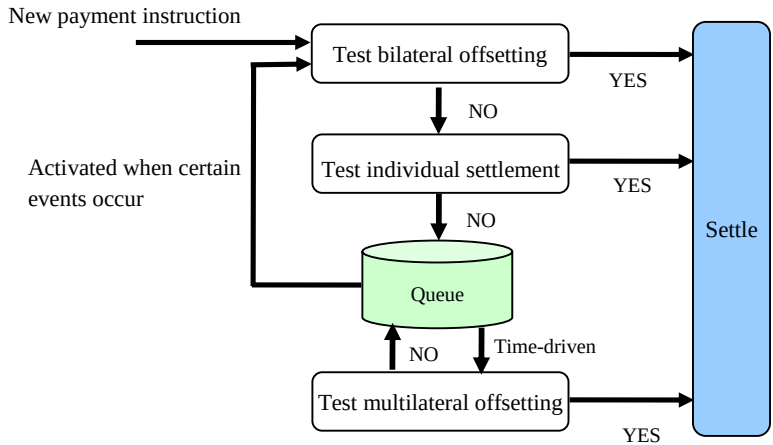


Figure: source: Bank of Japan.

Crucial: When large payments are received by the largest banks

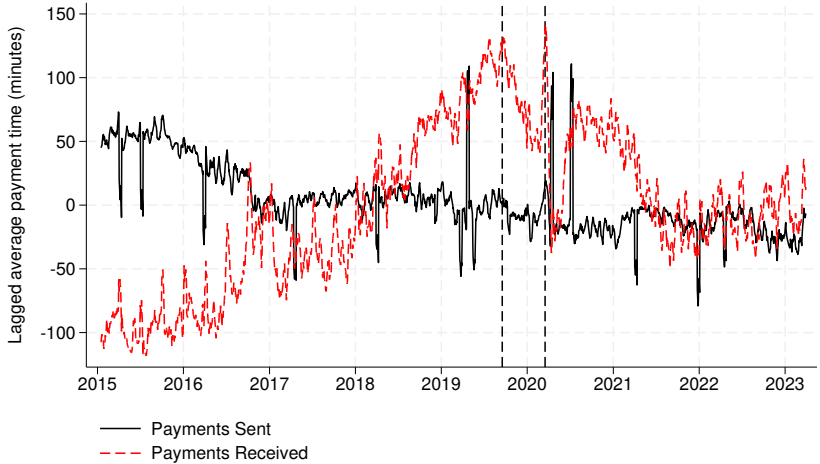


Figure: Time, relative to the sample mean, of the first half of large Fedwire payments, sent and received. Source: Copeland, Duffie, Wang, 2025.

Overdrafting significantly reduces the demand for reserves

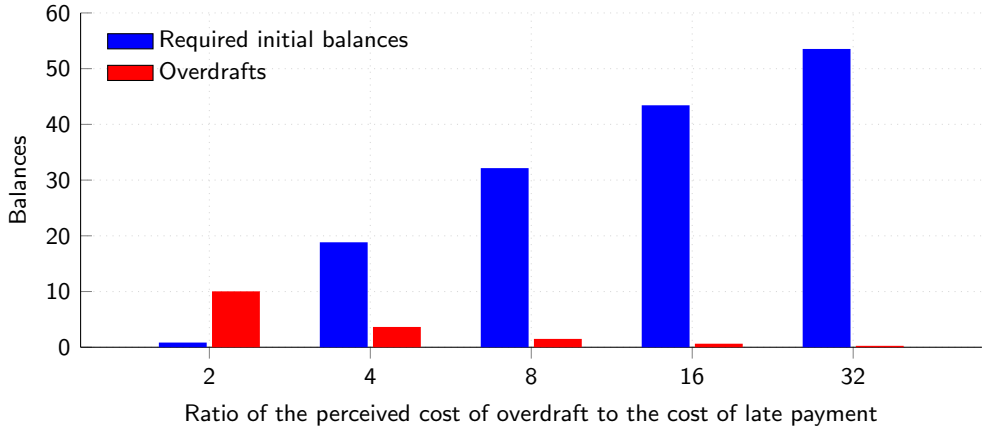


Figure: To achieve an expected fraction of 50% of early payments, the vertical axis shows required mean initial balances per bank (blue) and the mean chosen quantity of overdrafts per bank (red). Illustrative: based on the equilibrium payment timing model of Yang (2021).