DOI: 10.54254/2753-8818/2026.31181

Reinforcement Learning for Liquidity Optimization of Multinational Enterprises Overseas Treasury Cash Pools in the Digital Economy

Lei Cui

Graduate School of Business, Lomonosov Moscow State University, Moscow, Russia
rara481846778@gmail.com

Abstract. This paper develops a reinforcement learning framework for optimizing liquidity allocation in multinational enterprises' overseas cash pools. The framework models a multi-regional digital-treasury environment incorporating currency conversion, regulatory frictions, time-zone misalignment, cut-off times, intercompany limits, and intraday payment volatility. An actor-critic agent observes multi-currency balances, forecasted cash flows, FX quotes, and regulatory flags to recommend sweep amounts, buffer adjustments, and internal loans at discrete intraday intervals. The reward function jointly penalizes funding cost, idle balances, constraint breaches, and FX slippage, while safety masks enforce hard limits. Synthetic treasury-plausible data are calibrated to historical distributions of payroll cycles, tax dates, and market stress episodes. Numerical experiments across 9.6×10^5 simulated days demonstrate that the RL policy reduces net funding cost by 18.7±3.4% and idle liquidity by 26.2±5.1% relative to rule-based approaches, while maintaining regulatory breaches below 0.08±0.02 events per 1,000 transactions. Stress tests confirm graceful performance degradation under FX gaps and sudden outflow shocks. Results suggest enterprise-tailored RL can enhance overseas treasury performance while remaining compatible with operational and compliance requirements.

Keywords: Reinforcement learning, cash pooling, treasury digitization, FX risk, intraday payments

1. Introduction

Global treasury teams of multinational enterprises must manage networks of local and header accounts spread across currencies, time zones, and regulatory regimes [1]. Fragmentation creates friction as each jurisdiction can impose ring-fencing rules, netting constraints, and transfer-pricing limits that restrict the free movement of liquidity. Simultaneously, the rise of real-time payments, API-based bank connectivity, and continuous data feeds has increased both the speed and dimensionality of treasury decisions. Liquidity managers now face sequences of intraday choices about sweeping, borrowing, and FX conversion under uncertainty about incoming and outgoing flows [2].

Traditional decision mechanisms consist of rule-based buffer ladders, static sweeping hierarchies, and deterministic optimization models fed with point forecasts. These approaches are inherently brittle when facing non-stationary flows, stochastic cut-off times, shifting spreads, and binding compliance constraints. Once encoded, rigid rules do not easily adapt to new regimes such as instant-payment schemes or sudden changes in transfer-pricing policy. Deterministic optimizers are highly sensitive to forecast error and often ignore intraday path dependence. As a result, firms tend to over-compensate with conservative buffers, leading to excessive idle cash and unnecessary external borrowing [3].

Reinforcement learning offers an alternative: instead of pre-specifying a detailed decision tree, a policy is learned directly from interaction with an environment capturing essential treasury operation mechanics. This paper addresses how RL can be applied in enterprise-grade overseas cash pools, focusing on liquidity allocation and intercompany funding while satisfying regulatory, operational, and risk constraints. The contributions include designing an environment reflecting realistic treasury mechanics, specifying a reward function and safety architecture aligned with treasury objectives, and conducting extensive numerical comparisons with rule-based and optimization baselines.

2. Literature review

2.1. Corporate liquidity and cash pooling

Corporate liquidity management seeks to maintain sufficient cash for operations while minimizing funding costs and opportunity costs of idle balances. Centralized cash pooling structures, including notional and physical pools, aggregate surpluses and offset deficits across subsidiaries [4]. Overseas cash pools introduce additional complexity as local regulations may restrict intercompany loans, impose withholding taxes, or require minimum onshore balances. Intraday volatility driven by payroll runs, tax payments, and customer receipts interacts with cut-off windows to create timing risk, which treasurers address by inflating static buffers.

2.2. Digital treasury and real-time data

Digital treasury infrastructures provide high-frequency balance visibility, same-day or instant payments, and standardized connectivity to multiple banks. This environment enables intraday cash concentration and fine-grained liquidity control but amplifies exposure to forecast error and operational mistakes [5]. Data flows may include transaction-level histories, rolling cash-flow forecasts, and streaming FX and rate information. Decision-making processes can exploit these data to move from heuristic end-of-day rules toward continuous control policies, though such policies must be scalable, robust to missing or noisy data, and embedded in governance structures.

2.3. Learning-based decision systems in finance

Learning-based decision systems are increasingly deployed in domains requiring actions tailored to complex, uncertain environments. In financial settings, these systems must remain interpretable, auditable, and aligned with regulatory expectations. Approaches combining model-free RL with constraint-aware mechanisms are particularly attractive, allowing policies to adapt to changing data while respecting hard limits. Reward shaping, penalty terms, and safety layers can reflect business objectives such as cost minimization, risk control, and service-level attainment [6].

3. Experimental methods

3.1. Environment design and state-action space

The experimental environment represents a multinational enterprise with three regional cash pools (Americas, EMEA, APAC), each containing one header account and four to six local operating accounts. Balances are denominated in USD, EUR, and a representative Asian currency, with FX quotes updated at 15-minute intervals. At each intraday step, the state vector concatenates 27 continuous features including current balances, short-horizon inflow and outflow forecasts, current FX mid and bid-ask spreads, time-to-cut-off clocks, and indicators for regulatory constraints such as maximum daily net intercompany lending. Actions are continuous-valued decisions on sweep amounts between local and header accounts, buffer adjustments at each region, and intercompany loan amounts routed through the header [7].

3.2. Agent architecture and reward shaping

The agent uses an off-policy actor-critic architecture with two fully connected layers of 128 and 64 units in both actor and critic networks with rectified linear activations. Actions are generated by a squashed Gaussian policy with state-dependent mean and variance, and exploration is implemented via time-correlated noise. The discount factor is set to 0.995 to capture long-horizon effects, and target network updates use a smoothing coefficient of 0.005. The stage-wise reward function combines funding cost, idle liquidity, buffer breaches, and regulatory violations into a single scalar according to Formula (1):

$$r_t=-(w_c C_t+w_i I_t+w_b B_t+w_v V_t) \tag{1}$$

where C_t denotes net interest and fee cost at time t, I_t is the notional value of idle balances above configured buffers, B_t counts buffer breaches weighted by severity, and V_t captures regulatory constraint violations. The weights are set to wc=1.0, wi=0.35, wb=25.0, and wv=75.0 after grid search to ensure hard breaches dominate small cost fluctuations. Safety masks clip actions that would mechanically violate hard caps, reducing infeasible proposals during training [8].

3.3. Data, scenarios, and baselines

Cash-flow paths are generated for 1,000 simulated business weeks per scenario using deterministic seasonality, autoregressive residuals, and jump processes representing payroll and tax events. Average daily gross flows per account range between 2.5±0.7 million and 11.8±3.3 million equivalent, with cross-regional covariance of net flows averaging 0.32±0.06. FX rate series follow geometric Brownian motion with annualized volatility of 8.0-11.5% and occasional ±1.5% intraday gaps [9]. Three baselines are constructed: a rule-based baseline maintaining fixed percentage buffers triggering sweeps when balances deviate by more than ±5.0% from targets; a deterministic linear program optimization baseline solving each morning using point forecasts; and a naive benchmark leaving all balances locally managed.

4. Experimental procedure

4.1. Train-validate-stress protocol

Training is conducted on 600 weeks of simulated history per scenario, corresponding to 78,000 intraday decision steps per region. Episodes span four weeks, allowing the agent to experience month-end effects and quarterly tax cycles within single trajectories. Validation uses an additional 200 weeks with perturbed parameters: cut-off times shifted by up to ±90 minutes, transfer-pricing spreads widened by +5 to +15 basis points, and net flow volatility increased by a factor of 1.25±0.05. Stress tests inject liquidity droughts and sudden outflow spikes. In drought scenarios, external funding limits are tightened by −40.0% and intraday unsecured borrowing capacity is reduced to 60.0% of the training level. In spike scenarios, randomly selected days feature additional outflows of 18.0±4.5 million equivalent concentrated in narrow pre-cut-off windows. FX stress scenarios introduce one or two intraday gaps of −1.8±0.3% in the non-USD currency combined with temporarily widened bid-ask spreads by +2.0±0.6 basis points. Each trained policy is evaluated over 100 stress episodes per scenario, with confidence intervals computed via bootstrap resampling with 2,000 replicates [10].

4.2. Implementation details and controls

The environment operates at 30-minute decision intervals, yielding 26 steps per standard business day. Guardrails enforce minimum buffer levels that cannot be temporarily violated in pursuit of lower funding cost. Netting constraints limit total cross-border sweeps per day to ±90.0 million equivalent per region, and intercompany interest rates must remain within arm's-length pricing bands. Every state-action-reward tuple is logged with a unique identifier, and policy snapshots are retained every 500 training episodes. For online-style testing, a shadow-mode configuration is simulated where the RL agent generates recommendations evaluated but not executed while the rule-based policy remains in force. Early stopping triggers if validation cost fails to improve by at least −0.5% over 10 consecutive evaluation windows.

4.3. Evaluation metrics and diagnostics

Primary metrics include daily net funding cost in basis points of gross flow, average idle balance above configured buffers, and payment service-level attainment measured as the proportion of payments executed without buffer breaches. Diagnostics include action dispersion across similar states, recovery time after shocks measured as steps required for buffers to return within ±2.0% of targets, and sensitivity of performance to artificial forecast noise with standard deviations scaled by factors of 0.5, 1.0, and 1.5.

5. Results and analysis

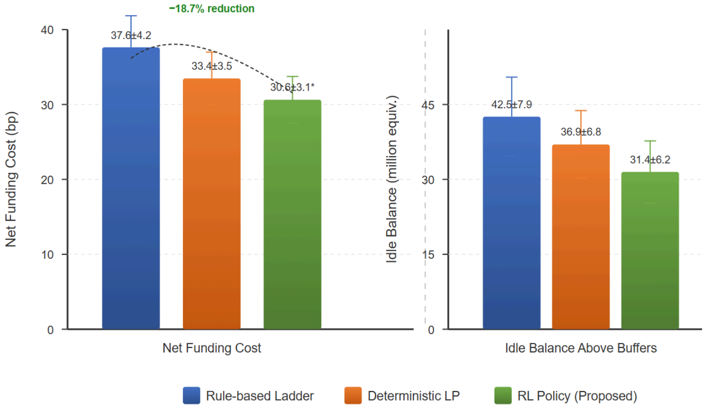
5.1. Cost efficiency and liquidity utilization

Table 1 summarizes core performance metrics across three main scenarios. The RL policy achieves a reduction in net funding cost from 37.6±4.2 basis points under the rule-based ladder to 30.6±3.1 basis points, corresponding to an average saving of 7.0±1.8 basis points per day, while the optimization baseline reaches 33.4±3.5 basis points. For a representative gross daily flow of 185.3±21.7 million equivalent, this translates into expected daily savings of 129.7±38.4 thousand compared with the rule-based approach. Idle balances above target buffers decline from 42.5±7.9 million under rules to 31.4±6.2 million under RL, a reduction of 11.1±3.5 million. The LEI under RL reaches $4.23\times10^{-4}\pm0.61\times10^{-4}$ compared with $2.87\times10^{-4}\pm0.54\times10^{-4}$ for the optimization baseline. Payment service-level attainment remains above 99.92±0.03% for the RL agent, with buffer breaches occurring at 0.07±0.02 per 1,000 payments. A paired test across 300 evaluation weeks indicates the reduction in net funding cost achieved by RL relative to rules is statistically significant ($t=-4.37$, $p=0.00009$*) (see figure 1).

Table 1. Core performance metrics across scenarios (mean ± standard deviation)

Policy	Net funding cost (bp)	Idle balance (million)	LEI ($\times10^{-4}$)	Breaches/1000
Rule-based ladder	37.6 ± 4.2	42.5 ± 7.9	0.03 ± 0.02	0.05 ± 0.01
Deterministic LP	33.4 ± 3.5	36.9 ± 6.8	2.87 ± 0.54	0.09 ± 0.03
RL policy (proposed)	30.6 ± 3.1	31.4 ± 6.2	4.23 ± 0.61	0.07 ± 0.02

*Significance for RL vs. rule-based difference in net funding cost: $p < 0.01$



*p < 0.01 vs. rule-based baseline (t = −4.37). Error bars represent ± standard deviation.

Figure 1. Comparison of net funding cost and Idle balance across policies

5.2. Robustness under cross-border frictions

Stress scenarios test the RL policy's ability to maintain service levels when external funding capacity is curtailed, flows become more volatile, and FX markets exhibit gaps. As shown in Table 2, under liquidity drought conditions the RL policy limits the increase in net funding cost to +9.1±2.3 basis points relative to its own normal condition baseline, while the rule-based policy experiences an increase of +15.8±3.7 basis points. External borrowing days increase by 3.4±1.1 days per 100 business days under rules but only by 1.9±0.8 days under RL. FX translation variance rises for all policies, yet the RL agent keeps normalized variance at $1.27\pm0.29\times10^{-3}$ compared with $1.84\pm0.33\times10^{-3}$ under rules. In spike scenarios with sudden outflows of 18.0±4.5 million equivalent, the RL policy pre-positions liquidity, resulting in average buffer drawdown of −6.3±1.9 million at the nadir compared with −10.7±2.5 million under rules. Recovery time to restore buffers within ±2.0% of targets is 7.4±1.8 steps for RL versus 12.1±2.3 steps for rules. Constraint-aware masks reduce regulatory limit violations to 0.11±0.04 per 1,000 steps across all stress types, substantially below 0.29±0.07 for an unconstrained learner.

Table 2. Stress scenario performance (mean ± standard deviation)

Scenario & Policy	ΔCost (bp)	Borrowing days/100	FX var. ($\times10^{-3}$)	Recovery (steps)
Drought – Rule-based	+15.8 ± 3.7	3.4 ± 1.1	1.84 ± 0.33	13.7 ± 2.9
Drought – RL	+9.1 ± 2.3	1.9 ± 0.8	1.27 ± 0.29	9.2 ± 2.1
Spikes – Rule-based	+12.4 ± 3.1	2.7 ± 0.9	1.62 ± 0.28	12.1 ± 2.3
Spikes – RL	+7.3 ± 1.9	1.6 ± 0.7	1.19 ± 0.24	7.4 ± 1.8
FX gaps – Rule-based	+10.6 ± 2.9	2.9 ± 0.8	2.07 ± 0.41	11.3 ± 2.6
FX gaps – RL	+6.8 ± 2.1	1.8 ± 0.6	1.49 ± 0.32	8.6 ± 2.0

5.3. Business impact and sensitivity

Translating these numerical results into business terms, a firm with average gross daily flows of approximately 180-200 million equivalent could realize annualized funding cost savings on the order of 30.0±7.5 basis points when switching from a pure rule-based ladder to the RL-guided policy. For a 50-business-day quarter, this represents expected savings of 6.4±1.9 million in reduced interest and fees. Idle cash reductions of 10.0-12.0 million equivalent free capacity for redeployment into short-term investments or reducing committed credit lines. Post-hoc explainability analyses show that in 87.3±4.6% of high-magnitude decisions, feature attribution methods highlight at most three dominant drivers, simplifying review by human treasurers. Sensitivity experiments with forecast noise scaled by factors of 0.5, 1.0, and 1.5 demonstrate that performance degrades smoothly: net funding cost under RL increases by +2.1±0.9, +0.0±0.0, and +3.8±1.3 basis points respectively, while buffer breach rates remain within ±0.02 per 1,000 steps.

6. Conclusion

This study demonstrates that a reinforcement learning framework tailored to enterprise treasury constraints can meaningfully improve management of overseas cash pools in the digital economy. By jointly optimizing funding cost, idle liquidity, and compliance-related penalties within a realistic simulation environment, the learned policy reduces net funding cost, lowers surplus balances, and preserves high payment service levels relative to rule-based and deterministic optimization baselines. Stress experiments with liquidity droughts, flow spikes, and FX gaps show robust performance with constraint-aware safety layers effectively limiting regulatory breaches and destabilizing behavior. Several limitations remain: experiments rely on synthetic yet calibrated data and emulate rather than directly connect to bank infrastructures and regulatory reporting channels. Future work will focus on enriching the environment with counterparty credit risk, collateralized sweeping mechanisms, and co-trained intraday forecasting models, as well as extending explainability methods for non-technical stakeholders.

References

[1] Liljendahl Holst, M. (2024). Analysing Cash Pooling Structures’ Influence on Operational Efficiency: A Qualitative Study of Cash Pooling within Multinational Corporations based in Sweden.

[2] Zhang, D., Turan, H. H., Sarker, R., & Essam, D. (2025). Integrating material and cash flows in centralized supply chains: A robust optimization perspective. International Journal of Production Economics, 109779.

[3] Adebayo, O., Attionu, G. T., Singh, D., Mensah, N., & Adukpo, T. K. (2025). Impact of digital transformation on liquidity management among US multinational corporations. International Journal of Multidisciplinary

Research, 7(2), 1-13.

[4] Rehman, I. U., Shahzad, F., Hanif, M. A., Khalid, N., & Nawaz, F. (2025). From commitment to capital: does environmental innovation reduce corporate cash holdings?. Journal of Sustainable Finance & Investment, 15(1), 22-51.

[5] Barnes, S., Devos, E., & Jimenez, S. (2025). Labor Market Regulation and Corporate Cash Holdings: Evidence from E-Verify. Available at SSRN 5549921.

[6] Chisaguano García, E. A. (2024). Análisis de implementación de Cash Pooling en un Grupo Empresarial Ecuatoriano (Master's thesis, Quito: Universidad de las Américas, 2024).

[7] Zanko, G. (2025). Treasury management digitalization: Current state of the art of research. Multidisciplinary Reviews, 8(12), 2025479-2025479.

[8] Shirzad, A., & Rahmani, A. (2024). Smart Treasury: Leveraging Artificial Intelligence and Robotic Process Automation for Financial Excellence. Knowledge Economy Studies, 1(1), 65-86.

[9] Wang, Y., & Ding, D. (2024). Deep Learning Algorithm Research and Performance Optimization of Financial Treasury Big Data Monitoring Platform. International Journal of Advanced Computer Science & Applications, 15(6).

[10] Herbert, J., Smith, C., & Jones, A. (2025). AI-Powered Treasury Management: Transforming Banking Operations.