

INTERNATIONAL MONETARY FUND

There Will Be Liquidity! But Will There Be Transmission?

Gregorio Impavido and Shuyu Wang

WP/26/109

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate.

The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

**2026
JUN**



WORKING PAPER

IMF Working Paper

Middle East and Central Asia Department

There Will Be Liquidity! But Will There Be Transmission?Prepared by **Gregorio Impavido and Shuyu Wang***

Authorized for distribution by Amina Lahreche

June 2026

IMF Working Papers describe research in progress by the author(s) and are published to elicit comments and to encourage debate. The views expressed in IMF Working Papers are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

ABSTRACT: This paper examines the impact of voluntary and involuntary excess liquidity on monetary policy effectiveness and inflation in Kazakhstan. Consistent with first-principles predictions, we find that voluntary liquidity held for precautionary motives is (i) negatively related to the opportunity cost of holding liquid assets and to the level of mandatory reserve requirements; (ii) positively related to average liquidity outflows, proxied by transactional demand for cash; and (iii) ambiguously affected by the magnitude and volatility of the business cycle. We further show that higher voluntary liquidity weakens monetary policy transmission and increases exchange rate pass-through to inflation. In contrast, involuntary liquidity hampers monetary policy effectiveness no matter its level and it increases average inflation primarily through its influence on the formation of inflation expectations. Overall, the results suggest that reforms aimed at reducing both forms of liquidity would enhance monetary policy effectiveness and, *ceteris paribus*, reduce inflation.

RECOMMENDED CITATION: Impavido, G. and S. Wang (2026) There Will Be Liquidity! But Will There Be Transmission? IMF working paper No WP/26/109

JEL Classification Numbers:	G21, G11, E52.
Keywords:	Liquidity risk; precautionary liquidity; involuntary liquidity; monetary policy; interest channel; credit channel; Kazakhstan.
Author's E-Mail Address:	gimpavido@imf.org , swang9@imf.org

* The views expressed in this paper are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management. The author(s) would like to thank Roberto Cardarelli, Amina Lahreche, Subir Lall, Giovanni Melina, and participants at seminars at the IMF for their invaluable comments. The usual caveat emptor applies.

WORKING PAPERS

There Will Be Liquidity! But Will There Be Transmission?

Prepared by Gregorio Impavido and Shuyu Wang¹

¹ The views expressed in this paper are those of the author(s) and do not necessarily represent the views of the IMF, its Executive Board, or IMF management. The author(s) would like to thank Roberto Cardarelli, Amina Lahreche, Subir Lall, Giovanni Melina, and participants at seminars at the IMF for their invaluable comments. The usual caveat emptor applies.

Table of Contents	Page
1 Introduction	1
2 Literature review	2
3 A simple model of reserve management	5
4 Select extensions to the base model	6
4.1 Mandatory reserve requirements	6
4.2 Liquidity risk	7
4.3 Business cycle	8
5 Estimating voluntary excess liquidity	9
6 Liquidity and monetary policy	12
7 Conclusions	15
References	18
Appendix I Figures and tables	22
Appendix II Data description and sources	30

List of Figures

1 Liquidity in Kazakhstan	22
2 Liquidity - cross country comparison	23
3 TONIA rate falling at the bottom of the interest band	23
4 Headline inflation	24
5 Excess liquidity	24
6 Required reserves	25
7 Currency in circulation	25
8 Lending rate	26
9 Actual and fitted excess liquidity values	28

List of Tables

1 Demand for liquidity - OLS regressions	26
2 Demand for liquidity - stationarity	27
3 Demand for liquidity - parameter stability	27
4 Demand for liquidity - diagnostic tests	27
5 Monetary policy - constrained OLS regression	28
6 Monetary policy - constrained threshold regressions	29
7 Monetary policy - linear restrictions across liquidity regions	29

1 Introduction

The National Bank of Kazakhstan (NBK) operates an inflation-targeting framework, using its policy rate—set within a symmetric corridor—as the main instrument to keep annual headline inflation near the 5 percent medium term target. As of end-December 2025, the policy rate stood at 18 percent, the Kazakhstan Stock Exchange (KASE) reverse repo facility, at the upper bound of the corridor, provided overnight liquidity at 19 percent, and the overnight deposit facility, at the lower bound of the corridor, absorbed excess liquidity at 17 percent. Since 2023, surplus liquidity has been primarily absorbed through daily 7-day deposit auctions conducted at the NBK’s initiative (Figure 1), on a full-allotment basis at the policy rate. Other liquidity absorption tools include notes at various maturities. In early 2026, the NBK expanded its short-term note issuance, capping 1-month notes at 1 trillion tenge and 3-month notes at 500 billion tenge.

Policy rate decisions and open market operations are aimed at steering short-term inter-bank rates around the base rate. In the absence of a reliable OTC-based inter-bank benchmark rate, the Tenge Over Night Index Average (TONIA) is currently the primary operational target of monetary policy for the NBK. It is a risk-free, weighted average, interest rate for one-day (overnight) repo transactions secured by government securities on the KASE. While the TONIA cannot directly affect variables such as the general price level, it can shape market expectations and financial conditions. In this manner, the operational target transmits policy signals via market interest rates, especially deposit and lending rates, thereby influencing price-setting behavior.

Monetary policy has been implemented in the context of ample liquidity. The buildup has primarily reflected unsterilized foreign exchange reserve accumulation associated with the sale of hydrocarbons and raw materials, substantial quasi-fiscal liquidity injections during and after the 2008 and 2014/15 banking crises (Mare et al. 2023), post COVID-19 pandemic support measures,¹ and targeted interventions to preserve financial stability at the onset of the war in Ukraine.² In particular, since 2011, the NBK has served as the primary purchaser of domestically produced gold. However, only since January 2025, has it adopted a “neutrality” principle for these purchases, sterilizing the associated tenge liquidity injections. Under this “neutrality” principle, gold purchases are offset in the following quarter by foreign exchange sales to sterilize the liquidity impact of earlier transactions and mitigate pressures on the local currency market. Purchase volumes are pre-announced on a monthly basis, operations are smoothed throughout the trading day, and the NBK acts as a price taker in the market. At end-2025, total liquidity reached about KZT 9.1 trillion including about KZT 6.6 trillion excess liquidity primarily absorbed through deposit-auctions and about KZT 2.6 trillion that banks kept in the correspondent accounts to comply with the required reserve regulation and meet precautionary liquidity needs (Figure 1).

¹Key measures, which resulted in an increase in deposits and therefore liquidity in the period 2020-22, included reducing risk-weighting ratios for SME and foreign currency loans, decreasing the capital conservation buffer and the liquidity coverage ratio to encourage loan restructuring.

²The NBK provided an interest subsidy on tenge deposits to fight dollarization in the context of weakening exchange rate which encouraged the accumulation of tenge liquidity.

Liquidity levels in Kazakhstan are high relative to other commodity-rich floating exchange rate economies. Cross-country comparisons are complicated by several factors, including differences in banks' funding structures, reserve requirement regimes, the state's footprint in the economy, caps on lending or deposit rates, and the extent to which governments or central banks fund the banking sector.³ Comparisons are also sensitive to where liquidity is recorded on balance sheets. Using comparable data across countries, liquidity as a share of bank deposits in Kazakhstan appears elevated (Figure 2).

Excess liquidity has weakened the link between the base rate and its operational target and the NBK is now re-calibrating the tools used to absorb liquidity. Since 2023, the TONIA rate has been frequently falling at the bottom of the interest rate corridor (Figure 3) suggesting a weakening of the interest rate channel of monetary policy and the need to re-calibrate liquidity absorption instruments. Inflation has been diverging from its target since 2019 (Figure 4). Starting in September 2025, with the aim to decrease its reliance on short-term liquidity tools, the NBK has started increasing the minimum reserve requirement and has plans to increase its reliance on central bank notes to absorb structural liquidity.

This paper examines the impact of voluntary and involuntary liquidity on monetary policy effectiveness and inflation in Kazakhstan. The remainder of the paper is organized as follows. Section (2) reviews the literature on banks' demand for liquidity and its implications for monetary policy transmission. Section (3) presents a simple liquidity risk management model to characterize the key determinants of voluntary liquidity demand. Section (4) considers select extensions of the base model to guide the selection of variables used in the following empirical strategy. This is implemented in two-stages: Section (5) estimates voluntary and involuntary liquidity demand, and Section (6) assesses their impact on monetary policy transmission and inflation. Section (7) concludes.

2 Literature review

The asset management literature on banks' demand for liquidity treats liquid assets as an insurance against early deposits withdrawal. In this literature, banks act as liquidity insurers for households facing idiosyncratic consumption shocks (Bryant 1980, 1981; Diamond and Dybvig 1983). When such shocks are imperfectly correlated, the required stock of liquid assets grows less than proportionally with total assets, allowing part of deposits to fund higher-yield, illiquid investments. The optimal size of a bank's liquidity buffer thus reflects a trade-off between the lower returns on liquid assets held for insurance purposes and the higher returns on illiquid assets (Baltensperger 1980; Santomero 1984).

This opportunity cost is not the only determinant of the demand for liquid assets. Banks also face a cost of not holding sufficient liquidity. When liquidity shortfalls arise, banks may be forced to obtain funding at punitive terms, such as through emergency market borrowing

³For example, in countries without a single treasury account, governments may hold cash balances in commercial banks, inflating measured liquidity.

or central bank standing facilities, implying a penalty associated with liquidity shortages. To these costs, one must add other non-negligible costs associated with portfolio adjustment (Baltensperger and Milde 1976, Agénor et al. 2004, and Stanhouse 1986). Banks may also need to invest in understanding depositors' types and behavior to minimize the risk of liquidity shortfalls. All these costs increase the bank's marginal cost of external finance in a moment of crisis (Calomiris et al. 2024). Using aggregate data for Thailand, Agénor et al. (2004) show that banks' liquidity holdings can also be positively related to the volatility of the money market rate, which, in turn, proxies the need for self-insurance. At the same time, Poole (1968) argues that a mean preserving spread in the distribution of deposit withdrawals should not necessarily result in higher reserve positions as the shock would leave banks with both excess and deficient reserves. This result may differ if banks were risk averse. Caprio and Honohan (1993) argue that credit demand or supply constraints also influence the build-up of excess liquidity on banks' balance sheet. For instance, unsatisfied demand in the goods market causes households to hold higher money balances that show up as higher deposits and excess reserves. This could be a structural phenomenon, like in the economies transitioning from central planning when goods rationing forced households to accumulate savings, or when soft budget constraints enable the State to crowd-out the private sector providing credit at below market clearing rates. It could also be a cyclical phenomenon when households expect constraints in the goods market to be relaxed. This generates a pent-up demand for future goods in the form of accumulated money balances, as it happened during the COVID-19 pandemic, rather than a reduction in the labor supply as it would be expected. Risk averse banks may also ration credit even if borrowers are willing to pay higher interest rates or are willing to borrow more. When banks are unable to differentiate the risk of the different borrowers, credit is rationed as higher lending rates may attract riskier borrowers (adverse selection) reducing portfolio profitability. Costly verification of the borrower's use of credit (moral hazard) could similarly lead to a backward bending supply curve as higher lending rates would increase the borrower's risk of default (Stiglitz and Weiss 1981; Akerlof 1970; Bester and Hellwig 1987; Williamson 1986, 1987; Townsend 1979; and Gale and Hellwig 1985). Other possible determinants of excess liquidity include: (i) large informal economy which generates bank deposits (and hence reserves) endogenously; (ii) high remittances that cause a build-up of deposits when households convert foreign currency into local currency; and (iii) unsterilized foreign exchange market interventions (Khemraj 2007).

Deposit withdrawals in excess of the expected household demand for liquidity needs generate negative externalities that cause banks to hold fewer liquid assets than what is socially optimal. The need to raise cash during a crisis lowers asset prices for other banks that might also need to sell. Individual banks do not account for this fire-sale externality leading them to hold too few liquid assets in aggregate. Alternatively, externalities can cause bank runs and transform idiosyncratic liquidity risk into systemic risk. Two key externalities have been identified as underlying causes of bank runs. In one class of models bank runs are caused by coordination failures among depositors (herding). When depositors observe early withdrawals by other depositors, they may interpret this as a signal of banks problems triggering a bank run. I.e., runs occur due to self-fulfilling nature of depositors' expectations concerning the behavior of their peers (Bryant 1980, 1981; Diamond and Dybvig 1983; Postlewaite and Vives 1987; Goldstein and Pauzner 2005; Rochet and Vives 2004; Freixas and Rochet 2008;

and Peck and Setayesh 2023). In the other class of models, bank runs occur as result of asymmetric information concerning bank fundamentals. In these models, depositors receive noisy signals about banks' profitability and a run can be triggered even in the good state of the world. By holding reserves, banks try to insulate themselves against the liquidity risk of a small number of "misinformed" early withdrawals in states of the world where the return on their assets is actually good (Chari and Jagannathan 1988; Jacklin and Bhattacharya 1988; Calomiris and Kahn 1991; Parlatore 2024). The presence of externalities associated with liquidity shocks justifies government intervention in the form of different regulatory responses and institutional arrangements to limit the probability of liquidity shortfalls, the possibility of contagion, and their costs. At the bank level, these include liquidity coverage, net stable funding, capital, and debt-equity convertibility regulations, among others. At the aggregate level these include mechanisms such as narrow banking, the lender of last resort, deposit insurance, and reserve requirements. These regulatory responses and institutional arrangements either reduce the need for self insurance by the individual bank or force them to endogenize the negative externality associated with liquidity shortfalls thereby forcing banks to hoard high-quality liquid assets, hence, reducing their demand for excess reserves.

The literature on the impact of excess liquidity on monetary policy effectiveness generally finds that excess liquidity negatively impacts the effectiveness of monetary policy with involuntary excess liquidity being particularly detrimental. Saxegaard (2006) investigates the relationship between excess liquidity and the effectiveness of monetary policy in Sub-Saharan Africa. It finds that an abundance of liquidity can diminish the impact of monetary policy measures on inflation and economic activity, primarily because excess liquidity leads to reduced interest rate sensitivity and can undermine policy transmission channels. In particular, it distinguishes between voluntary excess liquidity, held for precautionary motives, and involuntary excess liquidity. With voluntary excess liquidity, monetary policy would, in principle, be effective. A tightening (loosening) of the monetary stance would cause a deficit (surplus) in excess liquidity with respect to what is optimal for precautionary motives. To restore equilibrium, banks would contract (expand) credit to restore the optimal level of liquidity. When excess liquidity is involuntary banks have an additional constraint in adjusting credit in response to changes in the monetary stance. Bathaluddin et al. (2012) follows a similar approach and finds similar results for Indonesia. Girotti and Horny (2023) highlight how a tightening of the monetary stance increases excess liquidity remuneration and therefore, it reduces the opportunity cost of holding such reserves. Banks holding more excess liquidity experience a greater boost in revenue and net worth increasing their capacity to lend and therefore, reducing their incentive to transmit policy rate hikes to loan rates. In a similar vein, Agénor and Aynaoui (2010) argue that excess liquidity may cause deposit rate not to respond to a monetary contraction and to induce an easing of collateral requirements on borrowers (deterioration of underwriting standards) which, in turn, may translate into lower lending rates. Irrespective of the specific channel, an analysis of the impact of excess liquidity on the monetary policy transmission mechanism requires an understanding of the extent to which excess reserves are consistent with the commercial banks' demand for excess reserves for precautionary motives.

3 A simple model of reserve management

This section builds on Baltensperger (1980), especially section 2.1, Agénor et al. (2004), and section 8.2 of Freixas and Rochet (2008) to present a simple model of liquidity risk management for a bank.

Consider a representative bank with given liquid liabilities D (deposits) that invests in illiquid assets L (loans). The maturity mismatch between assets and liabilities gives rise to liquidity risk that the bank needs to manage. The way the bank manages this risk is to determine at the beginning of the period the optimal level R of interest bearing liquid assets (reserves) out of D .

Let r_D , r_L , and r denote the exogenous deposit, lending and reserve rates, respectively, and assume a risk neutral bank. At the end of the period, the net outflow of deposits is given by a random variable $x \sim \phi[0, D]$. If x does not exceed R (i.e., $x \leq R$), the bank has enough liquid assets to meet the net outflow of deposits. This happens with probability $\Phi(R) > 0$. If x exceeds R (i.e., $x > R$), the bank faces a liquidity shortage. This happens with probability $(1 - \Phi(R)) > 0$. In order to honor its liabilities, the bank needs to borrow the reserve deficiency $(x - R) > 0$ at penalty rate r_p yielding a penalty cost $r_p(x - R) > 0$ proportional to the borrowed amount. Under these assumptions, the bank's expected profit maximization problem is given by:

$$\max_R \quad E(\Pi) = r_L L + rR - r_D D - r_p \int_R^D (x - R) \phi(x) dx \quad (1a)$$

$$\text{s.t.} \quad L = D - R, \quad (1b)$$

$$x \sim \phi[0, D] \quad (1c)$$

After substituting the constraint (1b) into the objective function (1a), one can show that expected profits are a function of the bank's interest margin $(r_L - r_D)$ reduced by: (i) the opportunity cost $(r_L - r)$ of holding reserves,⁴ and (ii) by the expected cost of liquidity shortages $r_p \int_R^D (x - R) \phi(x) dx$.

The first order condition of the problem (4) is given by:

$$(r_L - r) = r_p [1 - \Phi(R)] \quad (2)$$

which says that the representative bank will hold reserves until the marginal opportunity cost of an additional unit of reserves $(r_L - r)$ is equal to the marginal reduction in expected cost of liquidity shortages $r_p (1 - \Phi(R))$.

⁴Notice that $(r_L - r) > 0$ as liquidity and credit risk premia dictate that $r < r_L$.

Hence, in equilibrium, the optimal level of reserves is given by:

$$R^* = \Phi^{-1} \left[\frac{r_p - (r_L - r)}{r_p} \right] \quad (3)$$

which is decreasing with the opportunity cost of holding reserves and increasing in the penalty rate.⁵

Notice that for $R^* > 0$ in equation (3), it is necessary that the penalty rate be larger than the opportunity cost of holding reserves (i.e., $r_p > r_L - r$). Now, the penalty rate r_p has often been interpreted as the discount rate for the bank to borrow the liquidity shortfall. If this were the case, it is unclear that the necessary condition for $R^* > 0$ be met. Essentially, banks are borrowing a liquid asset and, without additional information, the most that one can safely assume is that $r \leq r_p \leq r_L$, which is not sufficient for $R^* > 0$. One assumption that has been used to justify positive reserve holdings in equilibrium is to say that the marginal cost of borrowing liquidity r_p includes transaction costs and other inconveniences, such as the need for the bank to rearrange its portfolio at short notice, that could be very high (Baltensperger 1980). More intuitively, if the end-of-period penalty rate were to be lower than the beginning-of-period opportunity cost of holding reserves, the bank would have no incentives to hold reserves to start with.

4 Select extensions to the base model

The baseline model outlined above captures the core intuition of bank liquidity risk management but requires extension to improve realism and align with the determinants of liquid asset demand identified in Section (2) before empirical application in Section (5). This section examines the effects of mandatory reserve requirements—used as a proxy for government intervention—and liquidity risk, modeled as a function of unexpected fluctuations in cash demand and output, along with their associated volatility.

4.1 Mandatory reserve requirements

Under a contemporaneous reserve requirement regime, the bank must hold reserves at least equal to a fraction θ of deposits D .⁶ Total reserves can therefore be expressed as $R = \theta D + Z$, where Z denotes excess reserves.

⁵Using the inverse function theorem one can show that the derivative of the inverse of a strictly increasing cumulative distribution function (CDF) is also strictly positive.

⁶Contemporaneous reserves require banks to hold reserves against current deposits (same time period). They are intended to strengthen the linkage between bank reserves and the money supply to reduce short-term monetary fluctuations. Lagged reserves base requirements on past deposits (e.g., two weeks prior). They are intended to make it easier for banks to calculate and plan their required reserves with certainty, improving bank reserve management. Lagged reserve requirements are thought not to affect the demand for liquid assets by banks (Santomero 1984).

This requirement shifts the liquidity problem to the management of reserves in excess of the mandated level. It also lowers the critical value of the liquidity shortfall x beyond which a reserve deficiency and borrowing costs occur. Absent reserve requirements, this threshold equals beginning-of-period reserves R (Section (3)). With reserve requirements, the end-of-period total reserves must be greater or equal to the proportion θ of end-of-period deposits ($(R - x) \geq \theta(D - x)$) which implies that excess reserves must be $Z = (R - \theta D) = (1 - \theta)x$. The bank incurs borrowing penalty costs if the liquidity shock exceeds the threshold $x > Z/(1 - \theta) = \tilde{Z}$. The penalty costs are proportional to the reserve deficiency $(x - \tilde{Z})(1 - \theta)$.

With this modification, the bank's expected profits maximization problem becomes:

$$\max_Z E(\Pi) = r_L L + rR - r_D D - r_p(1 - \theta) \int_{\tilde{Z}}^D (x - \tilde{Z}) \phi(x) dx \quad (4a)$$

$$\text{s.t. } L = D - R, \quad (4b)$$

$$R = \theta D + Z, \quad (4c)$$

$$x \sim \phi[0, D] \quad (4d)$$

Hence, in equilibrium, the optimal level of excess reserves is given by:

$$Z^* = \Phi^{-1} \left[\frac{r_p - (r_L - r)}{r_p} \right] (1 - \theta) \quad (5)$$

which is decreasing with the opportunity cost of holding excess reserves $(r_L - r)$ and increasing in the penalty rate r_p , as in the case of no reserve requirements, but also a decreasing function of the mandatory reserve requirement rate θ .

4.2 Liquidity risk

The current model specification does not permit analysis of the effect of liquidity risk on the demand for excess reserves. To address this limitation, deposit withdrawals are assumed to follow a normal distribution,⁷ allowing liquidity risk to be characterized in terms of expected withdrawals and their volatility.

Let the deposit withdrawals $x \sim N(\mu, \sigma^2)$. As before, the bank incurs borrowing penalty costs if $x > Z/(1 - \theta) = \tilde{Z}$ with the reserve deficiency given by $(x - \tilde{Z})(1 - \theta)$. With a normal distribution, the probability of this happening is the same as the probability that $\tilde{x} = (x - \mu)/\sigma > (\tilde{Z} - \mu)/\sigma = \tilde{\tilde{Z}}$. The normalized reserve deficiency, when the liquidity shock exceeds this threshold, is now $\left[(\tilde{x}\sigma + \mu) - (\tilde{\tilde{Z}}\sigma + \mu) \right] (1 - \theta)$, or more simply $(\tilde{x} - \tilde{\tilde{Z}})\sigma(1 - \theta)$.

⁷This assumption is adopted for analytical tractability, despite permitting unbounded withdrawals.

With this modification, the bank's expected profits maximization problem becomes:

$$\max_Z E(\Pi) = r_L L + rR - r_D D - r_p \sigma (1 - \theta) \int_{\tilde{Z}}^{\infty} (\tilde{x} - \tilde{Z}) \phi(\tilde{x}) d\tilde{x} \quad (6a)$$

$$\text{s.t. } L = D - R, \quad (6b)$$

$$R = \theta D + Z, \quad (6c)$$

$$\tilde{x} \sim N[0, 1] \quad (6d)$$

where ϕ is now the standard normal density function.

The optimal level of excess reserves is given by:

$$Z^* = \left(\Phi^{-1} \left[\frac{r_p - (r_L - r)}{r_p} \right] \sigma + \mu \right) (1 - \theta) \quad (7)$$

which is decreasing with the opportunity cost of holding reserves ($r_L - r$), increasing in the penalty rate r_p , decreasing in the reserve requirement rate θ , increasing in the average deposit withdrawal μ , and increasing in the volatility of deposit withdrawal σ^2 .

The modification just made to our base model is not earth shattering: we simply say that higher liquidity risk, in the form of either higher average withdrawals μ or volatility σ , increases the demand for precautionary excess reserves.

4.3 Business cycle

Business and financial cycles are correlated (Borio et al. 2001 and Claessens et al. 2011). In the presence of financial frictions, such as costly state verification of idiosyncratic temporary shocks to entrepreneurs' net worth, the financial cycle shows a persistent impact on the business cycle (Bernanke and Gertler 1989). In other words, during economic expansions, loan demand rises, risk premiums fall, rates may ease (or banks ease standards) thus amplifying the business cycle in the upswing. During contractions (recessions), risk premia jump, loan standards tighten, rates often rise relative to safe assets, making borrowing harder, thus dragging recessions for longer than it would be otherwise.

We can introduce the business cycle and related changes in risk aversion for banks in our model by assuming that the demand for loan, the supply of deposits, and their interest rates are a function of the business cycle as follows:

$$\begin{aligned} L(r_L, Y) &= f(r_L(Y))Y \\ D(r_D, Y) &= g(r_D(Y))Y \end{aligned} \quad (8)$$

where Y is the business cycle, the demand for loans decreases with the lending rate ($\frac{df}{dr_L} < 0$), banks lower their underwriting standards to deliver a higher credit volume when the output gap opens ($\frac{dr_L}{dY} < 0$), and therefore the demand for loans increases when the output gap

opens $\left(\frac{dL}{dY} = \frac{\partial f}{\partial r_L} \frac{\partial r_L}{\partial Y} Y + f(r_L(Y)) > 0\right)$, the supply of deposits increases with the deposit rate $\left(\frac{dg}{dr_D} > 0\right)$, banks need to attract more funds to provide more loans when the output gap opens $\left(\frac{dr_D}{dY} > 0\right)$, and therefore the supply for deposits increases when the output gap opens $\left(\frac{dD}{dY} = \frac{\partial g}{\partial r_D} \frac{\partial r_D}{\partial Y} Y + g(r_D(Y)) > 0\right)$.

These considerations about the business cycle suggest that its net impact on liquid assets is ambiguous. On the one hand, a positive output gap reduces banks' demand for liquid assets as lending expands. On the other hand, in order to finance more loans, banks need to raise more deposits which, in turn, would lead to higher liquid assets (Acharya and Naqvi 2012), as this increases the cost of expected liquidity shortfalls.

The business cycle affects the demand of excess reserves also through the transactional demand for cash. A positive output gap increases the transactional demand for cash $\left(\frac{dC}{dY} > 0\right)$ and its volatility $\left(\frac{d\sigma_C}{d\sigma_Y} > 0\right)$. This, in turn leads banks to expect higher deposit withdrawals on average $\left(\frac{d\mu}{dY} = \frac{\partial \mu}{\partial C} \frac{dC}{dY} > 0\right)$ and higher volatility $\left(\frac{d\sigma}{d\sigma_Y} = \frac{\partial \sigma}{\partial \sigma_C} \frac{d\sigma_C}{d\sigma_Y} > 0\right)$ and therefore, lead them to increase their holdings of excess reserves.

5 Estimating voluntary excess liquidity

Sections (3) and (4) give us a rich enough model specification of the optimal demand for excess reserves⁸ that we can pass to the data with the following *a priori* impacts on excess reserves:

$$Z^* = g \left(\begin{matrix} r_L, r, r_p, \mu, \sigma, \theta, Y, \sigma_y, C, \sigma_C \\ - \quad + \quad + \quad + \quad + \quad - \quad \pm \quad \pm \quad + \quad + \end{matrix} \right) \quad (9)$$

We proceed by estimating equation (9) using OLS as follows:

$$\begin{aligned} \ln \left(\frac{EL}{D} \right)_t &= \alpha_0 + \alpha_1 \ln \left(\frac{EL}{D} \right)_{t-1} + \alpha_2 \ln \left(\frac{RR}{D} \right)_t + \alpha_3 \ln \left(\frac{C}{D} \right)_t \\ &+ \alpha_4 r_{L,t} + \alpha_5 d_{covid,t} + \alpha_6 d_{rus,t} + \epsilon_t \end{aligned} \quad (10)$$

where EL is the excess liquidity held by commercial banks as reported in the central bank monetary survey, D is total banking sector deposits, RR is the required reserves appearing in the central bank monetary survey and capturing the regulatory component of liquidity demand,⁹ C is the currency in circulation proxying household transactional demand for cash,

⁸Potential additional extensions to the model encompass the diversification of asset and liability portfolios, the interaction between deposits and capital, and the implications of information costs, deposit insurance, and lender of the last resort, among others. Each of these extensions would ultimately alter the opportunity cost associated with expected liquidity shortfalls, thereby influencing the demand for excess reserves in a manner analogous to the approach discussed in this section.

⁹This item includes the total deposits in the correspondent accounts.

r_L is the lending rate proxied by the consumer loans lending rate and representing the opportunity cost of holding excess reserves, d_{covid} and d_{rus} are step dummies controlling for structural breaks in liquidity associated with the COVID-19 and the Russia–Ukraine shocks, respectively, and ϵ is the usual error term.¹⁰

Figure (5) shows the ratio of excess liquidity to total bank deposits over the period December 2015 and December 2025. Excess liquidity increased sharply during the COVID-19 period and remained elevated thereafter. However, as predicted by our model in Section (4), starting in October 2025, it decreased rapidly when the central bank started increasing the reserve requirement. Figure (6) shows the reserve requirement ratio over the same period. Required reserves were eased after 2018 primarily to combat high dollarization rates, encourage the use of the tenge, and provide liquidity to a banking sector needing to repair its balance sheet after the 2014/15 banking crisis. In addition, reserve requirements declined markedly in 2023 due to a sharp decrease in FX reserve requirements. Figure (7) shows the currency-to-deposit ratio which increased gradually prior to the pandemic and moderated afterward. The decrease in currency in circulation in Kazakhstan after 2020 is primarily driven by a rapid government-supported transition toward a cashless economy, accelerated by the COVID-19 pandemic, technological advancements in banking, and the adoption of digital payments. While cash usage in transactions has fallen, the shift away from physical currency also stems from the growth of P2P (person-to-person) transfers used as substitutes for POS terminals, as well as the anticipation of a national digital currency (Digital Tenge). Finally, Figure (8) reports the lending rate, which declined gradually during 2016–2019 amid the NBK’s easing cycle, spiked temporarily in 2020 due to pandemic-related and exchange rate shocks, and remained elevated and more volatile from early 2021 onward, in line with policy rate hikes during the subsequent tightening cycle in response to rising inflation and geopolitical shocks.

Table (1) reports the OLS estimates based on monthly data from June 2016 to December 2025. All variables are demeaned and detrended¹¹ prior to estimation. Phillips–Perron tests reported in (Table 2) are broadly supportive of stationarity. The lagged dependent variable, $\ln(EL/D)_{t-1}$, is capturing persistence in excess liquidity and is positive and statistically significant. The estimated coefficient of about 0.42 implies that a 1 percent higher level of excess liquidity last month is associated with a 0.42 percent higher level this month, consistent with gradual adjustment in banks’ liquidity management. The coefficient on $\ln(RR/D)_t$ is negative and significant. A 1 percent increase in reserve requirement ratio is associated with roughly a 0.34 percent decline in excess liquidity, consistent with our model in Section (4), which predicts that once a larger share of reserves becomes mandatory, banks need less excess liquidity on top of that minimum. The currency-to-deposit ratio, $\ln(C/D)_t$, enters with a positive and significant coefficient. A 1 percent increase in cash demand relative to deposits is associated with a nearly proportional increase in excess liquidity, consistent with our model in Section (4). This suggests that stronger transactional demand for cash raises

¹⁰All variables are defined in [Appendix II](#).

¹¹This is done without taking into account the structural breaks in the series. Another valid alternative to this approach is to account for the mean and trend shifts associated with the structural breaks when demeaning and detrending. This would not require the use of the dummies in the regressions.

potential outflow risk and precautionary liquidity buffers. The lending rate, $r_{L,t}$, carries a negative and statistically significant coefficient. A 1 percent point increase in the lending rate is associated with a decline in excess liquidity of about 0.05 percent, consistent with our model in Section (4), that higher opportunity cost of holding liquid assets relative to lending are associated with lower excess liquidity. Both the COVID-19 and Russia–Ukraine war dummies are positive and statistically significant. The estimated coefficients imply that excess liquidity was approximately 40 percent and 24 percent higher during those periods, respectively.¹² This is also consistent with the structural upward shifts in EL/D observed in Figure (5).¹³

We test the stability of the estimated coefficients to assess whether changes in excess liquidity dynamics are driven by structural shifts or evolving economic conditions. We employ the SupF (max-Chow) test of Andrews (1993). Because the timing of a potential break is unknown, the test considers each admissible date (trimming the first and last 15 percent of observations) as a candidate break point. For each candidate break point, a step dummy is interacted with the variable, the corresponding Chow F statistic is computed, and the maximum value (the SupF statistic) is retained. The test is implemented variable by variable and results are reported in Table (3). Overall, the estimated coefficients are broadly stable over the sample period from June 2016 to December 2025. Only the coefficient for the reserve requirement shows weak evidence of a structural change, with the estimated break occurring around November 2018 and statistical significance at the 10 percent level. The remaining parameters exhibit no evidence of instability. Consistent with this, recursive CUSUM diagnostics do not reject overall parameter stability.¹⁴ These findings suggest that our model is relatively stable during crisis episodes affecting liquidity in levels rather than at the margin.

To assess the adequacy of the baseline specification, we conduct a standard battery of diagnostic tests (Table 4). The Breusch–Godfrey tests fail to reject the null of no serial correlation up to four lags, indicating that the residuals are not autocorrelated. The White test fails to reject homoskedasticity, suggesting that the error variance is stable over time (robust standard errors are nevertheless reported as a precaution). The Ramsey RESET test does not reject the null of correct functional form, providing no evidence of misspecification and omitted nonlinearities. Variance inflation factors are all below 3, with a mean of 2.19, well below conventional thresholds, suggesting no meaningful multicollinearity. Overall, the diagnostic results indicate that the specification is statistically well behaved over the sample period from June 2016 to December 2025.

¹²In a log–level regression $\ln(Y) = \beta X$, a one-unit increase in independent variable X is associated with an approximate (100β) percent change in dependent variable Y . This linear approximation is accurate when β is small (typically less than 0.15 in absolute value). For larger coefficients, the exact percentage change should be computed as $100(e^{\beta}-1)$ for greater precision.

¹³Although not reported, we also found that the output gap and its volatility have a non-significant impact on excess liquidity, consistent with our model in Section (4). Proxies for the penalty rate for liquidity shortages, the remuneration of liquid assets, and the volatility of the transactional demand for cash also have a statistically negligible impact on voluntary liquidity.

¹⁴The recursive CUSUM (cumulative sum) test is a general diagnostic of parameter stability under the null hypothesis of no structural break. The test statistic (0.2321) is below the 10, 5, and 1 percent critical values, indicating no evidence of overall instability

Figure (9) plots the actual and fitted values of the demeaned and detrended log of the excess liquidity-to-deposits ratio. The fitted series closely tracks the main cyclical movements, notably the sharp increase in 2020 and the subsequent gradual moderation. No persistent or systematic deviations between actual and fitted values are observed.

6 Liquidity and monetary policy

The previous section enables us to distinguish between voluntary excess reserves, using the fitted values of equation (10), and involuntary reserves in excess of the marginal liquidity benefit needed to offset the opportunity cost of holding them, using the residuals of equation (10).¹⁵

There are various situations that can explain the presence of involuntary excess reserves that, despite their lower return, are not used to extend more credit or invested in higher yielding securities. One obvious possibility is that securities alternative to reserves are not a perfect substitute. This, for instance, happens when the credit risk associated with alternative securities is too high, when the bond market is not developed (lack of yield curve, non-functioning inter-bank or money markets, illiquid secondary markets, etc...). Another possibility is that the economy is in a liquidity trap. In a standard liquidity trap, a monetary tightening by the central bank just leads to a decrease in excess reserves, possibly even below banks' prudential requirements. Another possibility is that prices or interest rates are regulated. In the presence of price/interest regulation, relative prices fail to provide the necessary signal to allocate savings in an efficient manner, possibly resulting in the build up of involuntary reserves. A similar outcome would materialize if credit is demand- or supply-constrained. When the government crowds out commercial banks through alternative sources of funding, banks are inevitably left with involuntary excess reserves. Another possibility is that when the banking sector is highly concentrated and uncompetitive, market power can result in lending rates that do not clear the market. Finally, banking sectors in resource-rich countries are likely to be characterized by excess involuntary liquidity, especially when large FX inflows needed to finance the extractive industry, or stemming from the sale of raw materials abroad, are not sterilized. In summary, there are compelling reasons to suggest that banks in Kazakhstan may hold excess reserves beyond what is needed to offset the marginal opportunity cost of holding them. These relate in particular to asymmetric information, lack of competition in the financial sector, underdeveloped bond markets, and large non-sterilized FX inflows.

The distinction between voluntary/precautionary and involuntary excess liquidity is not innocuous for the purpose of our analysis. When banks hold excess reserves only for precautionary purposes, one would still expect monetary policy to be effective. A loosening of monetary policy increasing excess liquidity above the level demanded by commercial banks for precautionary purposes would lead banks to expand lending by lowering the cost of borrow-

¹⁵Data limitations prevent us from explicitly modeling involuntary excess reserves as done for voluntary excess reserves.

ing or reducing the rationing of loans. Similarly, a contractionary monetary policy reducing liquidity below the level demanded by commercial banks for precautionary motives would lead banks to contract lending to maintain their desired level of excess liquidity. However, if the holdings of excess liquidity are involuntary in the sense that banks are unable to expand lending, changes in the monetary stance would only result in changes in the liquidity held by banks.

The analysis of the impact of liquidity on monetary policy effectiveness is carried out in two steps. In a first step we estimate a baseline constrained OLS regression for a modified Phillips curve along the lines proposed in Impavido (2025). This will be our benchmark model against which liquidity state-dependent models will be compared. In a second step, we use equation (10) to split total liquidity between a precautionary component, using its fitted values, and an involuntary component, using its residuals. We then assess the impact of monetary policy on inflation using threshold regressions in high and low voluntary and involuntary liquidity states.

In the first step of our empirical strategy we estimate the following baseline model using constrained OLS regression:

$$\pi_t = \beta_0 + \beta_1\pi_{t-1} + \beta_2E_t\pi_{t+1} + \beta_3ms_{t-2} + \beta_4xrt_{t-2} + \beta_5fs_{t-1} + u_t \quad (11)$$

where π_{t-1} and $E_t\pi_{t+1}$ are the backward- and forward-looking components of inflation expectations, respectively, ms is a proxy for the monetary stance, xrt is the log of the KZT/USD exchange rate, proxying external sources of inflation, and fs is a proxy for the fiscal stance. An increase in the exchange rate level amounts to a depreciation of the tenge, while an increase in the level of the monetary and fiscal stances amounts to a "tightening" of the same.¹⁶

Results over the period June 2016 - December 2025 are reported in Table (5). The parameters for the backward- and forward- looking components of inflation expectations are constrained to add up to one: i.e., $\hat{\beta}_1 + \hat{\beta}_2 = 1$. Results indicate that expectations are predominantly backward looking, as expected, given the high level of inflation in the country. A 1 percentage point tightening of the monetary stance yields about a 10 basis points decrease in headline inflation two periods ahead with the cumulative impact likely materializing only over a much larger horizon.¹⁷ A weakening of the tenge relative to the USD by 1 percent results in a low pass-through to inflation two months ahead of about 1 basis point. The estimated pass-through is lower than what is reported in Impavido (2025) as this model specification ignores the much higher indirect impact on inflation through changes in inflation expectations.¹⁸ Fiscal policy has the right sign but it has a statistically insignificant

¹⁶All variables are defined in [Appendix II](#).

¹⁷The NBK reports on its website that the full pass-through from base rate to inflation materializes, on average, between 12 and 18 months, as expected given the strong backward looking component of expectations.

¹⁸The NBK estimates a pass-through of about 1.4 percentage points for every 10 percent tenge depreciation over a period of three months.

impact on inflation and therefore, will be dropped in the remainder of the paper.¹⁹ The low value of the root mean squared error suggests a good model fit.

In the second step of our empirical strategy we estimate the same base model using threshold regressions. The estimation uses conditional least squares to estimate the parameters of the model in two regions of high or low liquidity. These two regions are identified by a liquidity threshold γ estimated by minimizing the residual sum of squares obtained from a grid search over all tentative thresholds. More formally:

$$\begin{aligned} \pi_t &= X_t\beta + Z_t\delta_1 + u_t & \text{if } -\infty < w_t \leq \gamma \\ \pi_t &= X_t\beta + Z_t\delta_2 + u_t & \text{if } \gamma < w_t \leq \infty \end{aligned} \tag{12}$$

where X contains the covariates with region-invariant parameters β , Z contains the covariates with region-specific coefficient vectors δ_1 and δ_2 , and w_t is a threshold variable.

Results for these threshold regressions are reported in Table 6, divided in four sections. The first section reports the estimated parameters of the region-invariant covariates. The second section reports the estimated parameters of the region-specific covariates in the *low liquidity region*, identified by the prefix R1. The third section reports the estimated parameters of the region-specific covariates in the *high liquidity region*, identified by the prefix R2. The fourth section reports summary statistics for each regression. Five different threshold regressions are presented:

Model 1 uses *involuntary liquidity* as the threshold variable w_t and it allows for only the impact of the monetary stance to be region-specific. Results for the region-invariant covariates are similar to our base model and therefore are not discussed here again. A change in the monetary stance has a similar impact on inflation as in our base model and not statistically different across liquidity regions as reported in Table (7).

Model 2 uses *voluntary liquidity* as the threshold variable w_t and it allows for only the impact of the monetary stance to be region-specific. A change in the monetary stance has an impact on inflation that is statistically stronger than in the baseline model when voluntary liquidity is low and weaker than in our baseline model when voluntary liquidity is high, as reported in Table (7). The joint assessment of Model 1 and 2 suggests that *involuntary liquidity* has a more detrimental impact on monetary policy effectiveness than *voluntary liquidity* respectively of its level.

Model 3 uses *involuntary liquidity* as the threshold variable w_t and it allows for both the impact of the monetary stance and exchange rate pass-through to be region-specific. A change in the monetary stance continues to have a similar impact on inflation as in our base

¹⁹This is likely due to various reasons. One reason is that the measure of fiscal stance used here fails to capture the impulses at the fiscal periphery for which data is not available at the frequency and length useful for this paper. Another reason is that we chose not to use a system estimation in order to conduct threshold analysis and addressed the issue of endogeneity by using lagged covariates. Indeed, Impavido (2025), using a SUR estimation, finds a strong impact of the fiscal stance on inflation directly, by supporting aggregate demand, and through the inflation expectation channel. Alternative specifications were tried using different lags for the covariates and found to yield similar qualitative results.

model and it is not statistically different across liquidity regions. Similar results apply to the exchange rate pass-through, as reported in Table (7).

Model 4 uses *voluntary liquidity* as the threshold variable w_t and it allows for both the impact of the monetary stance and exchange rate pass-through to be region-specific. A change in the monetary stance has an impact on inflation that is statistically stronger than in our base model when voluntary liquidity is low and lower than our base model when liquidity is high. The difference in impact across two liquidity regions is statistically significant at 5 percent level, as reported in Table (7). The exchange rate pass-through is stronger when voluntary liquidity is high and the different impact is statistically different from zero as reported in Table (7).

Involuntary liquidity may affect inflation through other channels. Model 5 uses *voluntary liquidity* as the threshold variable w_t and explores this possibility by adding a dummy variable for when involuntary liquidity is high as the only region-specific covariate. Results reported in Tables (6) and (7) suggest that average inflation is higher in the presence of high *involuntary liquidity*, particularly when *voluntary liquidity* is low, and inflation expectations are more backward looking and possibly de-anchored.

7 Conclusions

Monetary policy in Kazakhstan has been implemented in the context of large excess liquidity. This has been building up largely from unsterilized foreign exchange reserve accumulation, large quasi-fiscal liquidity injections to bail out banks during and after the last two banking crises in 2008 and 2014/15, respectively, injections associated with post-COVID support measures and interventions to safeguard financial stability through targeted support to commercial banks at the onset of the war in Ukraine. In addition, since 2011 the National Bank of Kazakhstan (NBK) has purchased gold from domestic producers to strengthen international reserves. However, only beginning in 2025 did the NBK start selling U.S. dollars in the domestic market to sterilize the tenge liquidity created by its gold purchases.

A simple first-principles model of banks' demand for voluntary (precautionary) liquidity predicts that voluntary liquidity is (i) negatively related to the opportunity cost of holding liquid assets and to the level of mandatory reserve requirements; (ii) positively related to the expected costs of liquidity shortages, the remuneration of liquid assets, average liquidity outflows and their volatility, and the transactional demand for cash and its volatility; and (iii) ambiguously affected by the level and volatility of the business cycle.

In the first stage of our empirical analysis, we find that the demand for voluntary liquidity in Kazakhstan is negatively associated with the reserve requirement, since higher mandatory reserves reduce banks' need to hold additional voluntary liquidity, and with the opportunity cost of holding reserves, as idle balances entail greater foregone income. By contrast, voluntary liquidity is positively associated with transactional demand for cash, reflecting the

higher probability of liquidity shortfalls when depositors' demand for cash increases.

We also find that the level and volatility of the business cycle have an ambiguous effect on optimal liquidity, owing to their contemporaneous positive influence on both deposit supply and loan demand. In contrast, proxies for the penalty cost of liquidity shortages, the remuneration of liquid assets, and the volatility of transactional cash demand have no statistically significant effect on voluntary liquidity.

The first stage of the empirical strategy also enables us to estimate the amount of liquidity involuntarily held by banks in excess of the level implied by optimal liquidity risk management. We define involuntary liquidity as the residual component driven by structural factors affecting loan demand and aggregate liquidity conditions, including the state's footprint in the lending market, the central bank's liquidity management and sterilization policies, and foreign exchange inflows.

In the second stage of our empirical strategy, we find that voluntary liquidity in Kazakhstan reduces the effectiveness of monetary policy in controlling inflation. This occurs through several channels. First, higher voluntary excess reserves lower the money multiplier (the ratio of money supply to the monetary base) so that for any unit of high-powered central bank money, commercial banks create less broad money. Second, the bank lending channel is weakened: higher voluntary excess reserves reduce the funds available for lending, thereby diminishing the marginal effect of central bank money on credit. Third, the interest rate channel is also impaired. By reducing the interest rate elasticity of credit supply, higher voluntary excess reserves limit the central bank's ability to stimulate aggregate demand through lower nominal rates. Given sticky prices, this mechanism reduces the effectiveness of monetary policy in increasing the real interest rates, as banks fail to fully transmit higher funding costs into lower lending, investment, or consumption.

We also find that exchange rate pass-through is stronger when voluntary liquidity is high. Since import prices typically respond one-to-one to exchange rate movements, albeit with a lag, depreciation shocks can place significant upward pressure on consumer prices through the imported component, increase inflation expectations, and raise the prices of non-tradable goods. Central banks typically respond to second-round effects of exchange rate shocks by tightening monetary policy. This action aims to anchor inflation expectations, prevent wage-price spirals, and counteract the inflationary pressure originating from higher import costs. However, since voluntary liquidity reduces the ability of monetary policy to control inflation by weakening the interest and credit channels as previously discussed, it also reduces the ability of monetary policy to anchor expectations and counteract the inflationary pressure originating from exchange rate shocks.

Finally, we find that involuntary liquidity is more detrimental for monetary policy effectiveness than voluntary liquidity and it is associated with less-anchored inflation expectations and higher average inflation. In other words, results suggests that involuntary liquidity influences inflation through channels distinct from those of voluntary liquidity. There are several potential channels through which involuntary liquidity can impact expectations. For instance, high involuntary liquidity may result from unsterilized liquidity injections, poten-

tially complicating the pursuit of immediate price stability. Under rational expectations, such a signal can lead agents to revise upward their beliefs about the central bank's future inflation tolerance, thereby increasing long-term inflation expectations. Similarly, although involuntary liquidity is not immediately convertible into credit, it may generate expectations that credit could be expanded rapidly if demand conditions change, thereby raising inflation expectations.

References

- Acharya, Viral and Hassan Naqvi (2012). “The seeds of a crisis: A theory of bank liquidity and risk taking over the business cycle”. In: *Journal of Financial Economics* 106 (2), pp. 349–366. DOI: <https://doi.org/10.1016/j.jfineco.2012.05.014>.
- Agénor, Pierre Richard and Karim El Aynaoui (2010). “Excess liquidity, bank pricing rules, and monetary policy”. English. In: *Journal of Banking and Finance* 34 (5), pp. 923–933. ISSN: 0378-4266. DOI: [10.1016/j.jbankfin.2009.10.003](https://doi.org/10.1016/j.jbankfin.2009.10.003).
- Agénor, Pierre Richard, Joshua Aizenman, and Alexander W. Hoffmaister (2004). “The credit crunch in East Asia: what can bank excess liquidity assets tell us”. In: *Journal of International Money and Finance* 23 (1), pp. 27–49. DOI: <https://doi.org/10.1016/j.jimonfin.2003.08.008>.
- Akerlof, George A. (1970). “The Market for "Lemons": Quality Uncertainty and the Market Mechanism”. In: *The Quarterly Journal of Economics* 84 (3), pp. 488–500. ISSN: 00335533, 15314650. URL: <http://www.jstor.org/stable/1879431>.
- Andrews, Donald W.K. (1993). “Tests for Parameter Instability and Structural Change With Unknown Change Point”. In: *Econometrica* 61 (4), pp. 821–856. DOI: <https://doi.org/10.2307/2951764>.
- Baltensperger, E. (1980). “Alternative approaches to the theory of banking firms”. In: *Journal of Monetary Economics* 6, pp. 1–37.
- Baltensperger, E. and H. Milde (1976). “Predictability of reserve demand, information costs and portfolio behavior of commercial banks”. In: *The Journal of Finance* 31 (3), pp. 835–843.
- Bathaluddin, M. Barik, Nur M. Adhi P., and Wahyu Ari Wibowo (2012). “The impact of excess liquidity on monetary policy”. In: *Bulletin of Monetary Economics and Banking* 14 (3), pp. 523–533. DOI: <https://doi.org/10.21098/bemp.v14i3.404>.
- Bernanke, Ben and Mark Gertler (1989). “Agency Costs, Net Worth, and Business Fluctuations”. In: *The American Economic Review* 79 (1), pp. 14–31.
- Bester, Helmut and Martin Hellwig (1987). “Moral Hazard and Equilibrium Credit Rationing: An Overview of the Issues”. In: *Agency Theory, Information, and Incentives*. Ed. by Günter Bamberg and Klaus Spremann. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 135–166. ISBN: 978-3-642-75060-1. DOI: [10.1007/978-3-642-75060-1_9](https://doi.org/10.1007/978-3-642-75060-1_9). URL: https://doi.org/10.1007/978-3-642-75060-1_9.
- Borio, Claudio, Craig Furfine, and Philip Lowe (2001). *Marrying the macro- and micro-prudential dimensions of financial stability*. BIS Paper 01. Basel, CH: Bank of International Settlements.

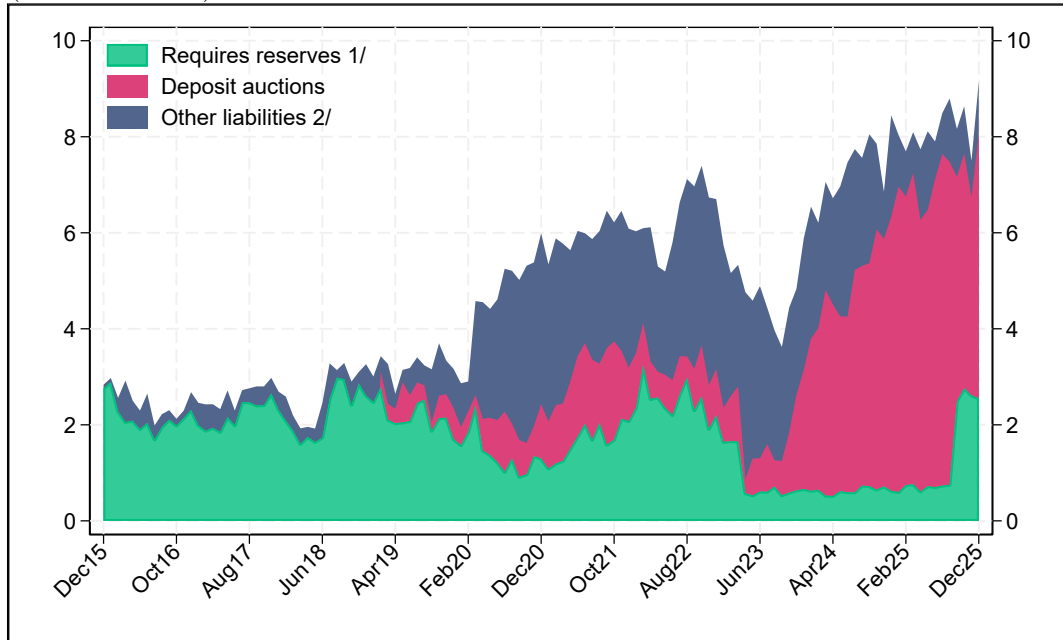
- Bryant, John (1980). “A model of reserves, bank runs, and deposit insurance”. In: *Journal of Banking and Finance* 4 (4), pp. 335–344. DOI: [https://doi.org/10.1016/0378-4266\(80\)90012-6](https://doi.org/10.1016/0378-4266(80)90012-6).
- (1981). “Bank collapse and depression”. In: *Journal of Money, Credit and Banking* 13 (4), pp. 454–464. DOI: <https://doi.org/10.2307/1991935>.
- Calomiris, Charles W. and Charles M. Kahn (1991). “The Role of Demandable Debt in Structuring Optimal Banking Arrangements”. In: *American Economic Review* 81 (3), pp. 497–513. URL: <https://ideas.repec.org/a/aea/aecrev/v81y1991i3p497-513.html>.
- Calomiris, Charles W., Madalen Castells-Jauregui, Florian Heider, and Marie Hoerova (2024). *A Theory of Bank Liquidity Requirements*. URL: https://www.tse-fr.eu/sites/default/files/TSE/documents/sem2024/bdf/heider_dgso_liquidity_regulation.pdf.
- Caprio, Gerard and Patrick Honohan (1993). “Excess liquidity and monetary overhangs”. In: *World Development* 21 (4), pp. 523–533. ISSN: 0305-750X. DOI: [https://doi.org/10.1016/0305-750X\(93\)90107-K](https://doi.org/10.1016/0305-750X(93)90107-K). URL: <https://www.sciencedirect.com/science/article/pii/0305750X9390107K>.
- Chari, Varadarajan and Ravi Jagannathan (1988). “Banking Panics, Information, and Rational Expectations Equilibrium”. In: *The Journal of Finance* 43 (3), pp. 749–761. DOI: [10.1111/j.1540-6261.1988.tb04606.x](https://doi.org/10.1111/j.1540-6261.1988.tb04606.x).
- Claessens, S., M.A. Kose, and M. Terrones (2011). *How do business cycle and financial cycle interact?* IMF Working Paper WP/11/88. Washington, DC: International Monetary Fund.
- Diamond, Douglas W. and Philip H. Dybvig (1983). “Bank runs, deposit insurance, and liquidity”. In: *Journal of Political Economy* 91 (3), pp. 401–419. DOI: <https://doi.org/10.1086/261155>.
- Freixas, X. and J.C. Rochet (2008). *Microeconomics of banking*. Second Edition. The MIT Press. ISBN: 978-0-262-06270-1.
- Gale, Douglas and Martin Hellwig (1985). “Incentive-Compatible Debt Contracts: The One-Period Problem”. In: *The Review of Economic Studies* 52 (4), pp. 647–663. URL: <https://ideas.repec.org/a/oup/restud/v52y1985i4p647-663..html>.
- Girotti, Mattia and Guillaume Horny (2023). “Monetary policy transmission through banks when liquidity is abundant but unevenly distributed”. In: *Finance Research Letters* (56), p. 104061. ISSN: 1544-6123. DOI: <https://doi.org/10.1016/j.frl.2023.104061>. URL: <https://www.sciencedirect.com/science/article/pii/S1544612323004336>.

- Goldstein, Itay and Ady Pauzner (2005). “Demand–Deposit Contracts and the Probability of Bank Runs”. In: *The Journal of Finance* 60 (3), pp. 1293–1327. DOI: <https://doi.org/10.1111/j.1540-6261.2005.00762.x>.
- Impavido, Gregorio (2025). *Inflation Determinants in Kazakhstan: A tale of (at least) two stories*. IMF Working Paper WP/25/210. Washington, DC: International Monetary Fund.
- Jacklin, Charles J. and Sudipto Bhattacharya (1988). “Distinguishing Panics and Information-Based Bank Runs: Welfare and Policy Implications”. In: *Journal of Political Economy* 96 (3), pp. 568–592. DOI: [10.1086/261552](https://doi.org/10.1086/261552). URL: <https://ideas.repec.org/a/ucp/jpolec/v96y1988i3p568-92.html>.
- Khemraj, Tarron (2007). “Monetary policy and excess liquidity: the case of Guyana”. In: *Social and Economic Studies* 56 (3), pp. 101–127. URL: <https://mpra.ub.uni-muenchen.de/53126/>.
- Mare, Davide S., Martin Melecky, and Hanna Murina (2023). *Bank Bailouts and Fiscal Contingent Liabilities*. World Bank Policy Research Working Paper 10273. Washington, DC: The World Bank Group.
- Parlatore, Cecilia (2024). “Transparency and bank runs”. In: *Journal of Financial Intermediation* 81 (3), pp. 497–513. DOI: <https://doi.org/10.1016/j.jfi.2024.101120>. URL: <https://www.sciencedirect.com/science/article/pii/S1042957324000482>.
- Peck, James and Abolfazi Setayesh (2023). “Bank Runs and the Optimality of Limited Banking”. In: *Review of Economic Dynamics* 47, pp. 100–110. DOI: [10.1016/j.red.2022.01.003](https://doi.org/10.1016/j.red.2022.01.003). URL: <https://ideas.repec.org/a/red/issued/21-90.html>.
- Phillips, P. C. B. and P. Perron (1988). “Testing for a unit root in time series regression”. In: *Biometrika* 75, pp. 335–346. DOI: <https://doi.org/10.2307/23361>.
- Poole, William (1968). “Commercial Bank Reserve Management in a Stochastic Model: Implications for Monetary Policy”. In: *The Journal of Finance* 23 (5), pp. 769–791. DOI: <https://doi.org/10.2307/2325906>.
- Postlewaite, Andrew and Xavier Vives (1987). “Bank Runs as an Equilibrium Phenomenon”. In: *Journal of Political Economy* 95 (3), pp. 485–491. DOI: <http://dx.doi.org/10.1086/261468>.
- Rochet, Jean-Charles and Xavier Vives (2004). “Coordination Failures and the Lender of Last Resort: Was Bagehot Right After All?”. In: *Journal of the European Economic Association* 2 (6), pp. 1116–1147. DOI: <https://doi.org/10.1162/1542476042813850>. URL: <https://ideas.repec.org/a/tpr/jeurec/v2y2004i6p1116-1147.html>.
- Santomero, Anthony M. (1984). “Modeling the Banking Firm: A Survey”. In: *Journal of Money, Credit and Banking* 16 (4), pp. 576–602.

- Saxegaard, Magnus (2006). *Excess Liquidity and Effectiveness of Monetary Policy: Evidence from Sub-Saharan Africa*. IMF Working Paper WP/06/115. Washington, DC: International Monetary Fund.
- Stanhouse, B. (1986). “Commercial bank portfolio behavior and endogenous uncertainty”. In: *The Journal of Finance* 41 (5), pp. 1103–1114.
- Stiglitz, Joseph E. and Andrew Weiss (1981). “Credit Rationing in Markets with Imperfect Information”. In: *The American Economic Review* 71 (3), pp. 393–410. ISSN: 00028282. URL: <http://www.jstor.org/stable/1802787>.
- Townsend, Robert M (1979). “Optimal contracts and competitive markets with costly state verification”. In: *Journal of Economic Theory* 21 (2), pp. 265–293. ISSN: 0022-0531. DOI: [https://doi.org/10.1016/0022-0531\(79\)90031-0](https://doi.org/10.1016/0022-0531(79)90031-0). URL: <https://www.sciencedirect.com/science/article/pii/0022053179900310>.
- Williamson, Stephen D. (1986). “Costly monitoring, financial intermediation, and equilibrium credit rationing”. In: *Journal of Monetary Economics* 18 (2), pp. 159–179. ISSN: 0304-3932. DOI: [https://doi.org/10.1016/0304-3932\(86\)90074-7](https://doi.org/10.1016/0304-3932(86)90074-7). URL: <https://www.sciencedirect.com/science/article/pii/0304393286900747>.
- (1987). “Costly Monitoring, Loan Contracts, and Equilibrium Credit Rationing”. In: *The Quarterly Journal of Economics* 102 (1), pp. 135–145. ISSN: 00335533, 15314650. URL: <http://www.jstor.org/stable/1884684>.

Appendix I Figures and tables

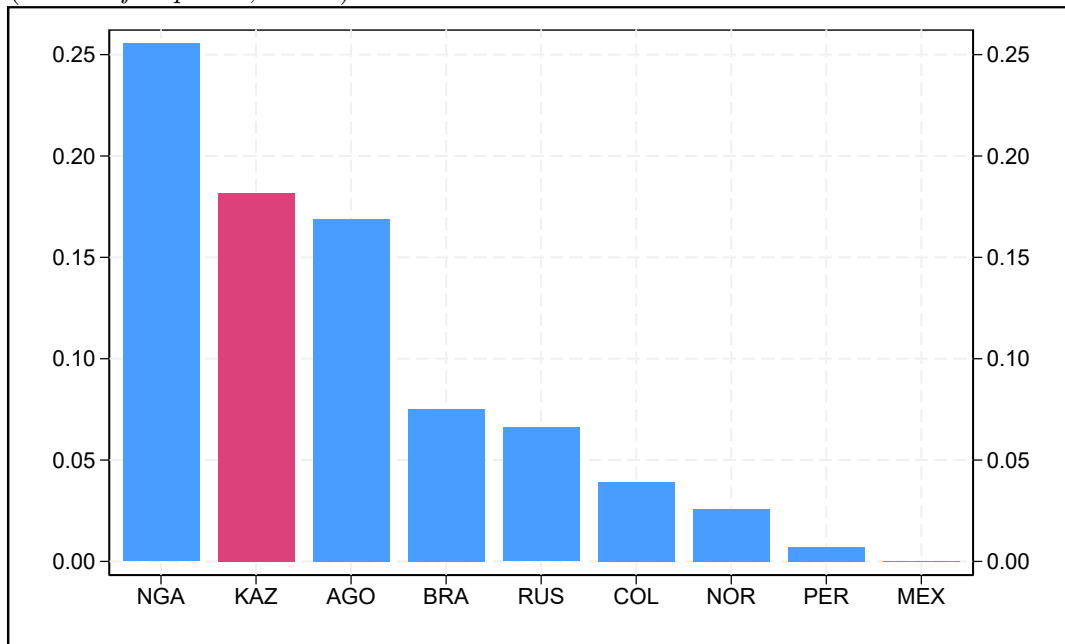
Figure 1: Liquidity in Kazakhstan
(KZT Trillion)



Source: NBK monetary survey and authors' calculations.

Notes: 1/ In correspondent accounts. 2/ Including excess reserves in correspondent accounts, standing liquidity facilities, open market operations, and other facilities.

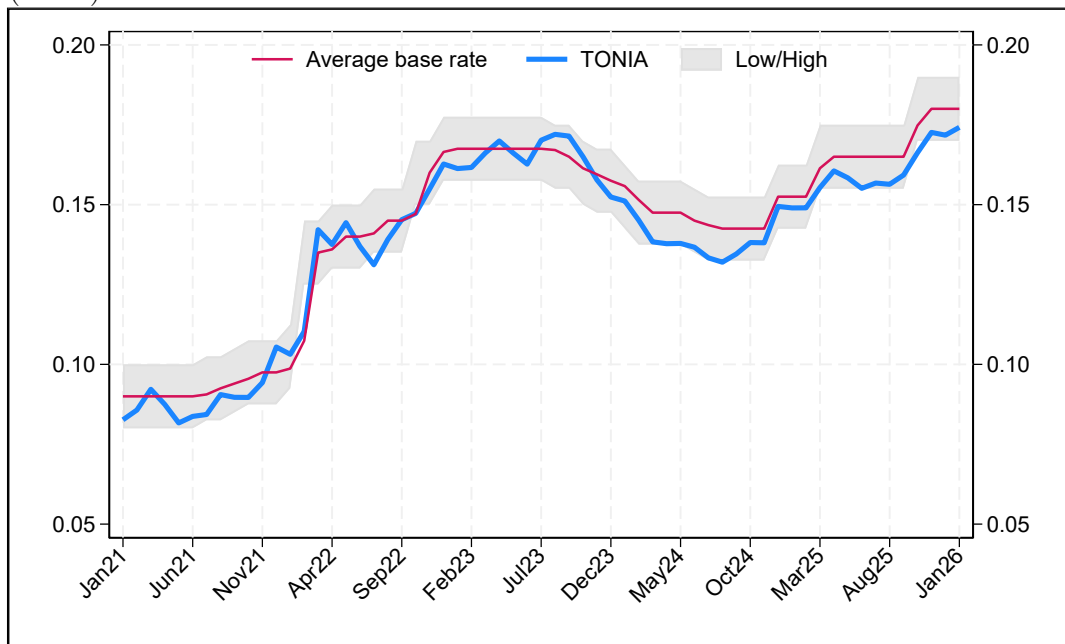
Figure 2: Liquidity - cross country comparison^{1/2/3/}
(Share of deposits, units)



Source: CB and ODC monetary surveys and authors' calculations.

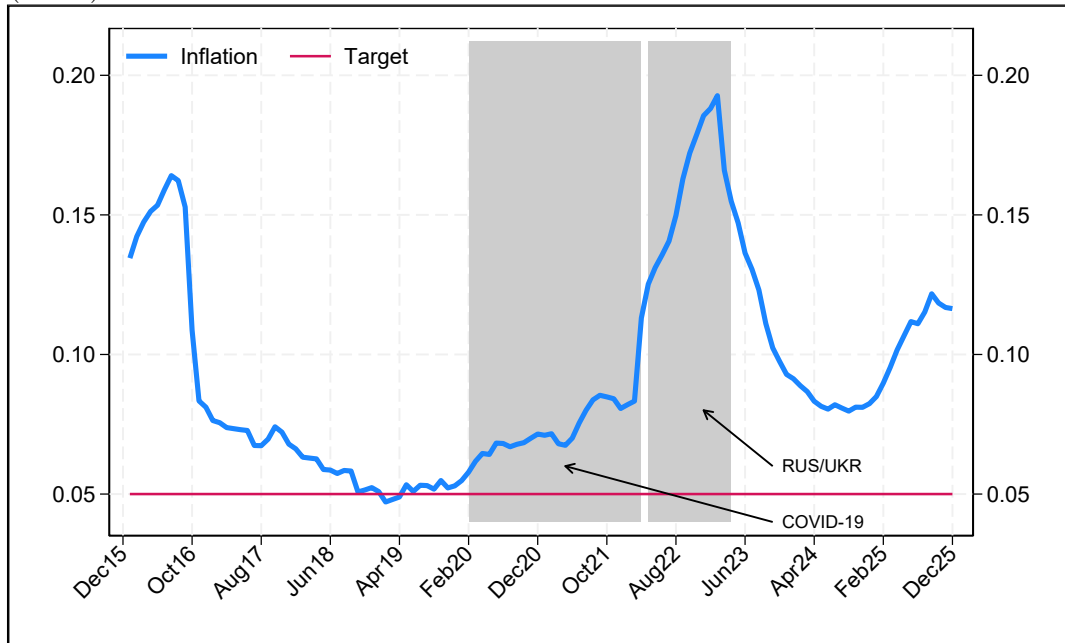
Notes: 1/ Comparable data across countries. 2/ Dec. 2025 or latest available. 3/ Peru Aug. 2025, Colombia and Norway Nov. 2025, Nigeria Sept. 2023, Russia Jan. 2022.

Figure 3: TONIA rate falling at the bottom of the interest band
(units)



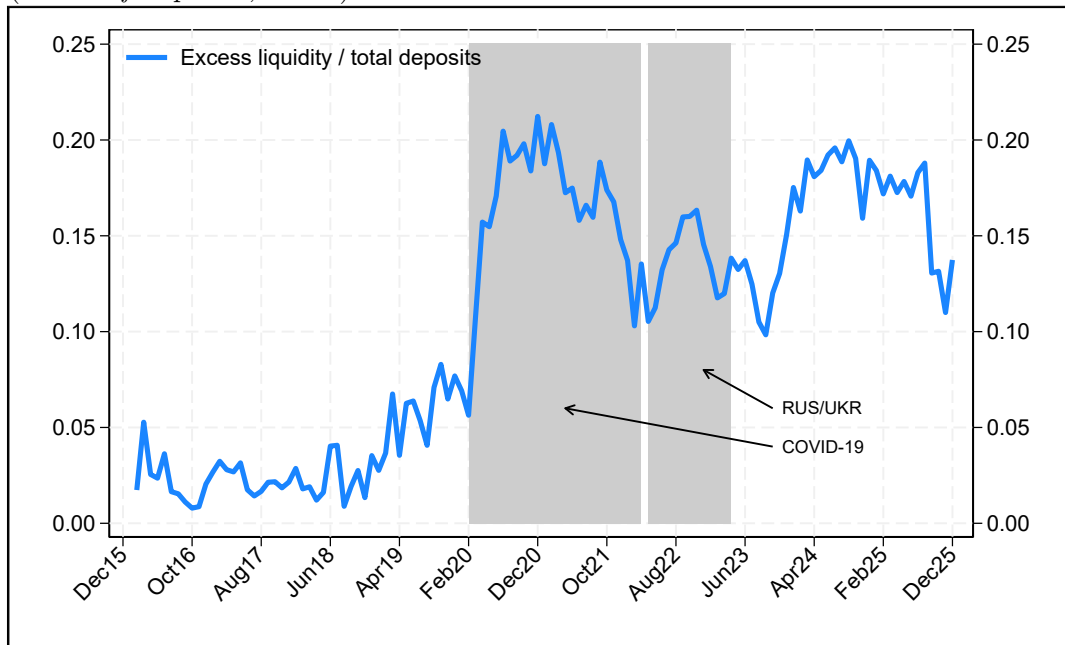
Source: Haver and authors' calculations.

Figure 4: Headline inflation
(units)



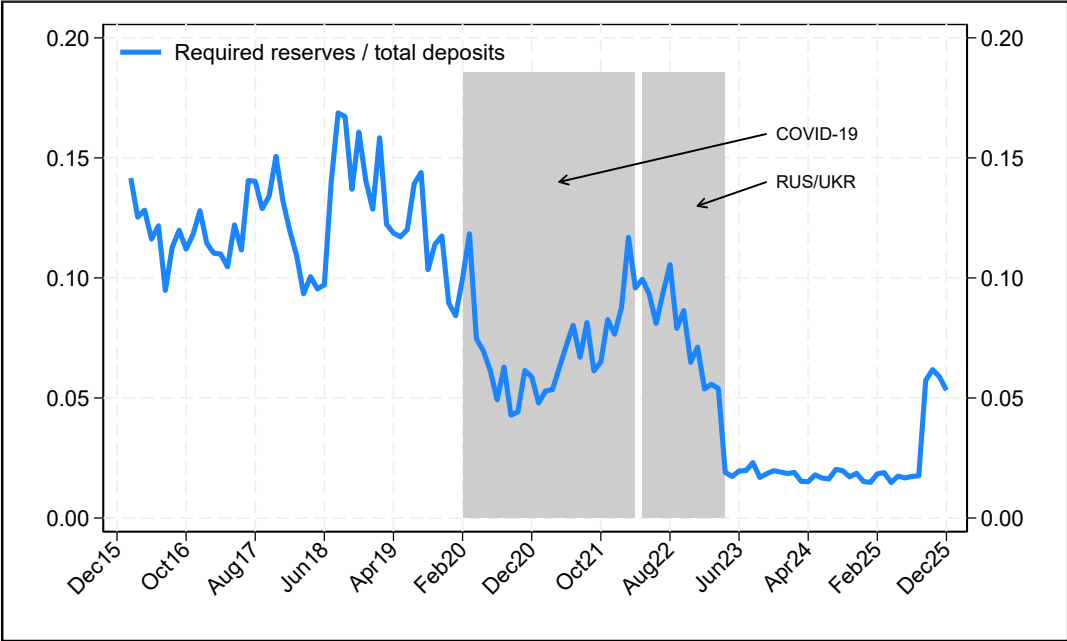
Source: Haver and authors' calculations.

Figure 5: Excess liquidity
(Share of deposits, units)



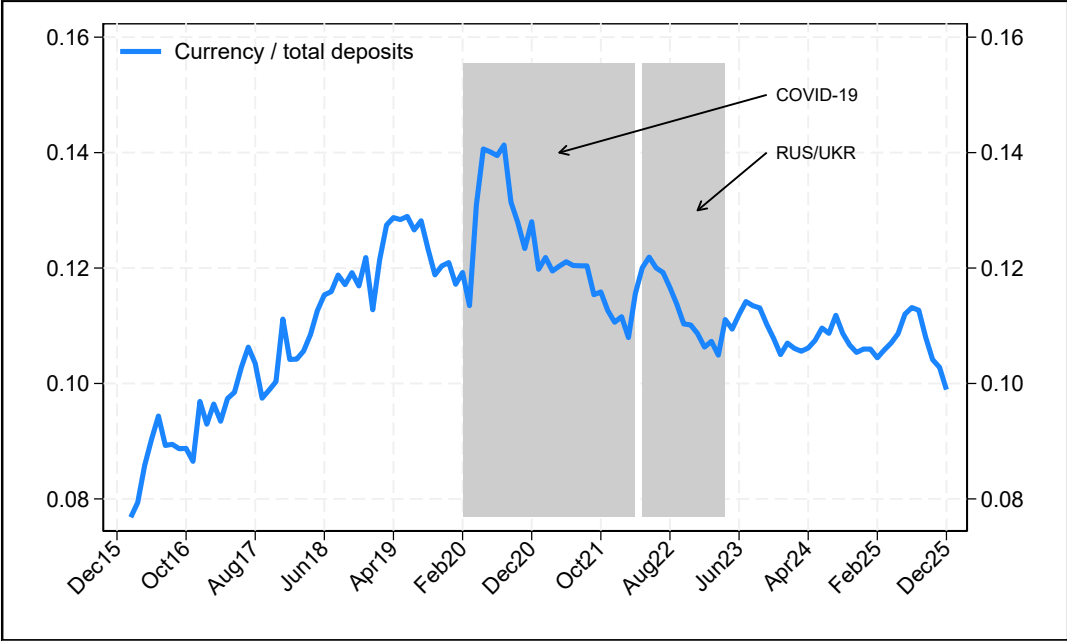
Source: Haver and authors' calculations.

Figure 6: Required reserves
(Share of deposits, units)



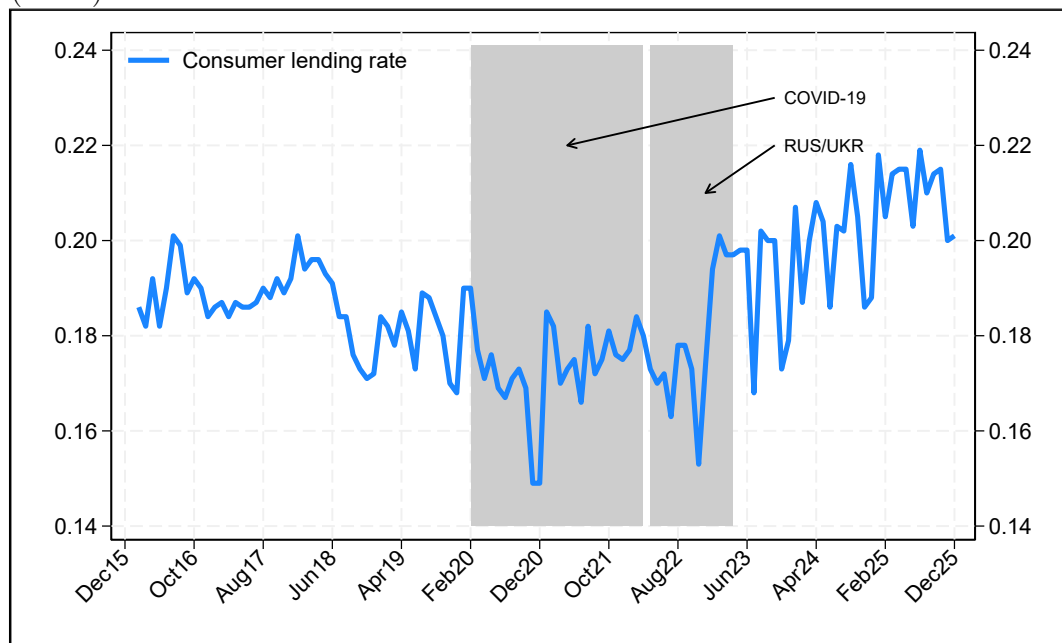
Source: Haver and authors' calculations.

Figure 7: Currency in circulation
(Share of deposits, units)



Source: Haver and authors' calculations.

Figure 8: Lending rate
(units)



Source: Haver and authors' calculations.

Table 1: Demand for liquidity - OLS regressions^{1/;2/;3/}

Variable	Model 1	Model 2
$\ln(E/D)_{t-1}$	0.423*** (0.079)	0.423*** (0.100)
$\ln(RR/D)_t$	-0.343*** (0.097)	-0.343*** (0.110)
$\ln(C/D)_t$	0.979*** (0.334)	0.979*** (0.345)
$r_{L,t}$	-0.047* (0.027)	-0.047** (0.021)
$d_{covid,t}$	0.396*** (0.103)	0.396*** (0.114)
$d_{rus,t}$	0.244** (0.105)	0.244** (0.096)
c	-0.127*** (0.038)	-0.127*** (0.049)
Obs	115	115
\bar{R}^2	0.786	0.786

Source: Authors' calculations.

1/ Estimation sample spans from June 2016 – December 2025. 2/ Model 2 reports heteroskedasticity-robust standard errors. 3/ *** = Pval < 0.01. ** = 0.01 <= Pval < 0.05. * = 0.05 <= Pval < 0.10.

Table 2: Demand for liquidity - stationarity^{1/, 2/, 3/}

Variable	Z(t)	P-Value
$\ln(E/D)$	-2.975 ^{**}	0.037
$\ln(RR/D)$	-2.585 [*]	0.096
$\ln(C/D)$	-2.746 [*]	0.066
fs	-2.894 ^{**}	0.046
r_L	-4.483 ^{***}	0.000
$ogap$	-3.499 ^{***}	0.008
r_{tonia}	-2.816 [*]	0.056
r_{pol}	-2.838 [*]	0.053

Source: Authors' calculations.

Notes: 1/ Phillips and Perron (1988) test with H0: unit root without drift. 2/ Critical values are -2.579 (10 percent), -2.889 (5 percent), and -3.504 (1 percent). 3/ *** = Pval < 0.01. ** = 0.01 <= Pval < 0.05. * = 0.05 <= Pval < 0.10.

Table 3: Demand for liquidity - parameter stability^{1/, 2/, 3/}

Variable	Sup F	Break date
$\ln(E/D)$	3.685	Nov-18
$\ln(RR/D)$	8.787 [*]	
$\ln(C/D)$	2.215	
r_L	3.649	

Source: Authors' calculations.

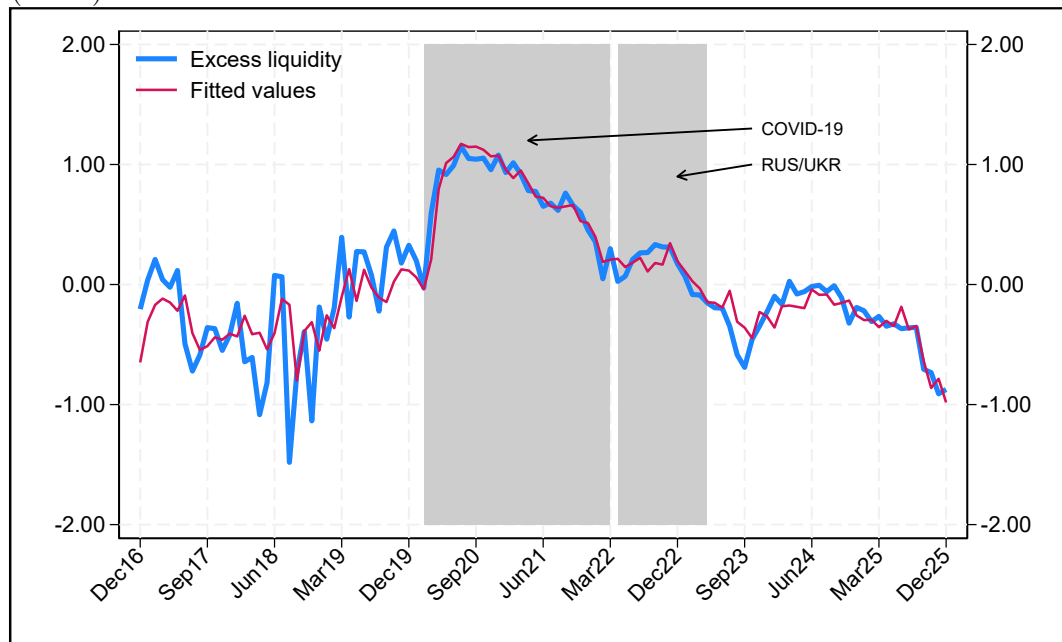
1/ SupF (max-Chow) statistics are computed following Andrews (1993) for single-parameter structural break tests with 15 percent trimming ($p = 1$). For each variable, the reported statistic is the maximum Chow F statistic across admissible break points. 2/ Critical values for $p = 1$ and 15 percent trimming are 7.17 (10 percent), 8.85 (5 percent), and 12.35 (1 percent). 3/ *** = Pval < 0.01. ** = 0.01 <= Pval < 0.05. * = 0.05 <= Pval < 0.10.

Table 4: Demand for liquidity - diagnostic tests

Test	Distrib.	Value	P-Value
White heterosk.	$\chi^2(24)$	24.150	0.453
BG LM (lags 1-4)	$\chi^2(4)$	6.228	0.179
RESET	F(3,105)	1.150	0.332
VIF		2.190	

Source: Authors' calculations.

Figure 9: Actual and fitted excess liquidity values
(units)



Source: Authors' calculations.

Table 5: Monetary policy effectiveness - constrained OLS regression^{1/}

Variable	$\hat{\beta}$	s.e.	t	P-Value	Sig. ^{2/}
π_{t-1}	0.947	0.022	43.696	0.000	***
$E_t\pi_{t+1}$	0.053	0.022	2.446	0.016	**
ms_{t-2}	-0.114	0.021	-5.314	0.000	***
xrt_{t-2}	0.010	0.004	2.268	0.025	**
fst_{t-2}	0.012	0.015	0.791	0.431	
c	-0.058	0.026	-2.223	0.028	**
Obs	115				
Root MSE	0.006				

Source: Authors' calculations.

1/ Estimation sample spans from June 2016 – December 2025. 2/ *** = Pval < 0.01. ** = 0.01 <= Pval < 0.05. * = 0.05 <= Pval < 0.10.

Table 6: Monetary policy effectiveness - constrained threshold regression^{1/, 2/}

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
π_{t-1}	0.951*** (0.021)	0.968*** (0.022)	0.949*** (0.021)	0.941*** (0.021)	0.994*** (0.022)
$E_t\pi_{t+1}$	0.049** (0.021)	0.032 (0.022)	0.051** (0.021)	0.059*** (0.021)	0.006 (0.022)
xrt_{t-2}	0.007* (0.004)	0.018*** (0.006)	0.004 (0.004)
ms_{t-2}	-0.107*** (0.019)
R1. ms_{t-2}	-0.090*** (0.024)	-0.264*** (0.064)	-0.090*** (0.024)	-0.173*** (0.027)
R1. xrt_{t-2}	0.009** (0.004)	-0.001 (0.005)
R1. $invol$	0.018*** (0.004)
R1. c	-0.045* (0.023)	-0.098** (0.038)	-0.054** (0.025)	0.010 (0.030)	-0.038 (0.024)
R2. ms_{t-2}	-0.150*** (0.042)	-0.104*** (0.023)	-0.151*** (0.042)	-0.076** (0.033)
R2. xrt_{t-2}	0.005 (0.012)	0.038*** (0.013)
R2. $invol$	0.002* (0.001)
R2. c	-0.039* (0.024)	-0.108*** (0.039)	0.036 (0.070)	-0.229*** (0.079)	-0.024 (0.024)
Obs	115	115	115	115	115
AIC	-1158.164	-1162.339	-1157.527	-1166.705	-1175.647
γ	0.014	0.113	0.014	0.140	0.022
RSS	0.004	0.004	0.004	0.004	0.004

Source: Authors' calculations.

1/ Estimation sample spans from June 2016 – December 2025. 2/ *** = Pval < 0.01. ** = 0.01 <= Pval < 0.05. * = 0.05 <= Pval < 0.10.

Table 7: Monetary policy - linear restrictions across liquidity regions

	Model 1	Model 2	Model 3	Model 4	Model 5
H0: R1. ms_{t-2} = R2. ms_{t-2}	1.573	5.374	1.641	5.186	...
$\chi^2(1)$	0.210	0.020	0.200	0.023	...
H0: R1. xrt_{t-2} = R2. xrt_{t-2}	1.275	7.639	...
$\chi^2(1)$	0.259	0.006	...
H0: R1. $invol$ = R2. $invol$	17.701
$\chi^2(1)$	0.000

Source: Authors' calculations.

Appendix II Data description and sources

We build a dataset of monthly observations over the 2016m6-2025m12 interval. The dataset combines measures of required reserves, excess liquidity, banking sector aggregates, liquidity costs, inflation and inflation expectations, and indicators of the monetary and fiscal stances. All variables are defined as follows:

- π_t **Headline inflation.** Headline inflation is calculated as the 12-month difference in the log of the seasonally adjusted monthly CPI index. The CPI series is seasonally adjusted by Haver and published with the code S916PC.
- ms_t **Monetary stance.** Defined as the average ex-ante real policy rate with equally weighted inflation projections for the following twelve months. The monthly average policy rate is published by the Central Bank and reported by Haver with code N916RTAV.
- fs_t **Fiscal stance.** Defined as the difference between the logs of the GG 12 month running sum of revenues and expenditure reported by Haver with codes Y916FTRX and Y916FTEX, respectively.
- epi_t **Inflation expectations.** As published by the Central Bank and reported by Haver with code N916VMPE.
- xrt_t **US dollar exchange rate.** Defined as the monthly EOP value of tenge per unit of US dollars and reported by Haver as N916XUSE.
- RR_t **Required reserves.** Reported in the central bank monetary survey as the sum of codes 91614CRN.R...{Z} and 91614CRF.R...{Z}.
- EL_t **Excess liquidity.** Reported in the central bank monetary survey as "other liabilities" aggregating codes 91614CSN.R...{Z}, 91614CUN.R...{Z}, 91614CWN.R...{Z}, 91616DGF.R...{Z}, 91614CUF.R...{Z}, and 91614CWF.R...{Z}.
- D_t **Deposits.** Defined as the sum of demand, time, and savings deposits held at deposit money banks. Data are from the Banking System Monetary Survey compiled by the NBK and accessed via Haver with code N916FMRD.
- C_t **Currency in circulation.** Defined as currency held outside deposit money banks. Data are obtained from the NBK and accessed via Haver with code N916FMC.
- $r_{L,t}$ **Lending rate.** Proxied by the household lending rate. Data are from the NBK and accessed via Haver with code N916RLH.
- $d_{covid,t}$ **COVID-19 dummy.** Step dummy equal to one during the period March 2020 - March 2022 and zero otherwise.
- $d_{rus,t}$ **Russia-Ukraine war dummy.** Step dummy equal to one during the period April 2022 - March 2023 and zero otherwise.



PUBLICATIONS

There Will Be Liquidity! But Will There Be Transmission?
Working Paper No. WP/2026/109