The double-faced demand for bank reserves and its implications[1]

Claudio Borio, Head of the Monetary and Economic Department of the BIS
Piti Disyatat, Assistant Governor, Monetary Policy Group, Bank of Thailand
Andreas Schrimpf, Head of Financial Markets, Monetary and Economic Department, BIS

As they have engaged in quantitative easing, central banks have fundamentally adjusted their approach to influencing short-term market rates, shifting from scarce reserve systems to abundant reserve systems. This column argues that the shift has fundamentally altered the properties of the demand for reserves. While in a scarce reserve system reserves only serve as settlement medium, in an abundant reserve system they also become a store of value. This makes the demand for reserves partly endogenous to its supply and has material implications for the predictability of that demand, for how reserves influence the constellation of market rates, and for any transition back to a scarce reserve system. These issues are especially important as central banks evaluate the merits of the two systems and how best to implement policy.

Since the Great Financial Crisis (GFC), the analysis of the day-to-day implementation of monetary policy has attracted increasing attention. Operating frameworks have important implications for issues that pre-crisis were largely irrelevant or, given the small size of many central bank balance sheets, mostly relegated to technical discussions. These include helicopter money (e.g. Reichlin et al. 2013, Borio et al. 2016), the functioning of short-term money markets (Markets Committee 2019), central bank profits and government deficits (Bell et al. 2023), the transmission mechanism (Bigio and Sannikov 2021), and financial stability (Greenwood et al. 2016). The reason underlying this renewed interest is the shift in monetary policy operating frameworks, from scarce reserve systems (SRS) to abundant reserve systems (ARS) – sometimes also known as corridor and floor systems, respectively (e.g. Borio 1997, Borio and Disyatat 2010, Bindseil 2016, Cap et al. 2020).

The contours of these monetary policy operating systems are well known. In an SRS, the central bank limits the supply of reserves to keep the overnight rate above the deposit facility rate; in an ARS, it increases the supply to push the rate to that floor.[2]

What is less appreciated is that the function reserves play, and hence *the relevant properties of the demand for them, differs fundamentally across the two systems*. This regime dependence effectively means that demand is not independent of supply – i.e. it is partly *endogenous* – since it is the supply of reserves that determines which regime is in place.[3] In addition, it has implications for the stability of the demand for reserves and for how the systems influence market rates more broadly. Understanding this is especially important to inform the central bank reviews of operating frameworks under way.

---

[2] Here, we use the term 'abundant' to cover all systems where banks hold more reserves than needed for settlement purposes and where the opportunity cost for doing so is negligible for all or part of this excess. This does not differentiate between 'ample', 'abundant' or 'tiered' systems. On ample versus abundant, see for example Logan (2023) and Afonso et al. (2023); as an example of tiered, see Maechler and Moser (2022).

[3] Importantly, the endogeneity of reserves demand we highlight here is *different from the one that could reflect ratcheting or hysteresis effects* – the possibility that, the longer the ARS is in place, the higher the demand for reserves becomes. This phenomenon was observed, for example, in Norway and was one reason why the central bank took one step back towards an SRS (Central Bank of Norway 2021).

The regime dependence of the demand for reserves

In what sense is the demand for reserves regime-dependent? In an SRS, the demand is exclusively for settlement balances – for reserves as a *settlement medium*; in an ARS, the demand is also for portfolio purposes, most importantly at the margin – for reserves as a *store of value (i.e. investment purposes)*.

The key difference between the two types of demand is the relevance of the yields on alternative assets. These yields are effectively irrelevant for holdings of settlement balances, since the penalty for not meeting the payment commitments is prohibitive. By contrast, they matter a great deal for holdings of store-of-value balances, for which risk–return considerations are paramount. Put differently, *the demand for reserves as a settlement medium is not interest rate-sensitive.*[4] This is true regardless of whether additional mandatory holdings, in the form of reserve requirements[5] or any supervisory liquidity requirements, are present; in such cases, the penalty for not meeting the minima is also very high. By contrast, when the central bank satiates the market with excess reserves, beyond settlement needs, compelling banks to hold them, the risk–return profile of the reserves *must be sufficiently attractive relative to that of other assets* so that banks are willing to hold them. Relative yields become an additional variable in the demand function (also see Lopez-Salido and Vissing-Jorgensen 2023).

Australia: Reserve balances are independent of the policy rate   Graph 1

Lhs: —— Cash rate   – – Lending facility   Rhs: ▬ Reserve balances
     —— Cash rate target   – – Deposit facility
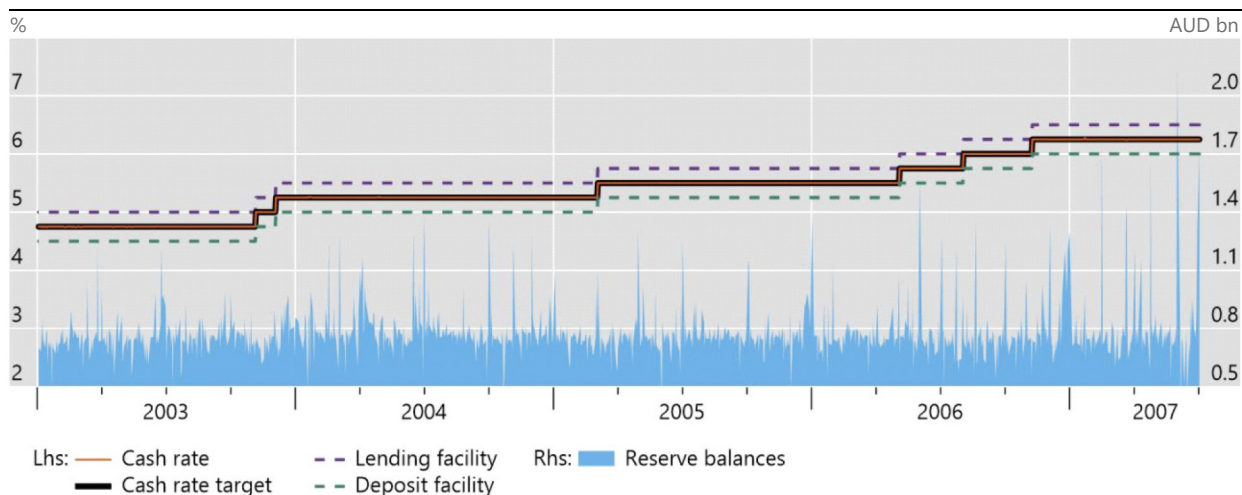
Source: Reserve Bank of Australia.

That the demand for *settlement* balances is unresponsive to interest rates is illustrated most clearly in the case of Australia, where the central bank does not impose reserve requirements (Graph 1). We see how, in the period before the GFC, the amount of settlement balances is tiny and independent of the overnight rate. To anchor the overnight rate, the central bank provides a clear *signal*[6] as to where it wants the rate

---

[4]   In principle, given optimising behaviour and uncertainty, it is possible to introduce some interest elasticity, not least given the possibility of borrowing from the central bank at a penalty rate; see Bech and Monnet (2013) for a review of modelling approaches. In practice, however, that elasticity is not exploitable to control the overnight rate in the absence of some sort of signalling. See Borio (1997) for a survey of central banks' signalling approaches.

[5]   Averaging provisions in reserve requirements ('maintenance periods') introduce 'indifference regions', as banks become more tolerant of shortfalls and excesses over the relevant time intervals. The tolerance declines as the end of the maintenance period nears, since this reduces the banks' room for manoeuvre to compensate for deviations from the average. Hence the typical spikes in overnight rates at the end of the periods.

[6]   The signal is so powerful because the central bank has a monopoly over the supply of reserves and can always peg the rate, if it wants, by standing ready to supply or withdraw as much reserves as needed at the target rate.

to be and makes sure that it meets the inelastic demand through liquidity management operations in order to keep rate volatility in check.

The implications of the regime dependence

At least two significant implications follow from this analysis: one concerns the transmission mechanism of policy; the other the transition from ARS to SRS.

The *transmission mechanism* is very different in the two systems. In an SRS, the amount of reserves is *neutral* with respect to market rates other than the overnight rate (operating target): the transmission operates *exclusively* from the anchored overnight rate to other rates. In turn, the overnight rate is insensitive to other yields in the system. By contrast, in an ARS, exogenous changes in quantities (i.e. the supply of reserves) will directly influence a broad array of market rates, as banks evaluate reserves relative to other competing stores of value. Put differently, with the return on reserves fixed at the remuneration rate, *other rates must adjust* to accommodate variations in the supply of reserves. The demand for bank reserves, in turn, will be sensitive to changes in those rates. *As a result, the amount demanded will shift along with the amount supplied as those yields adjust.*

New Zealand: Reserve balances push down the FX swap rate   Graph 2
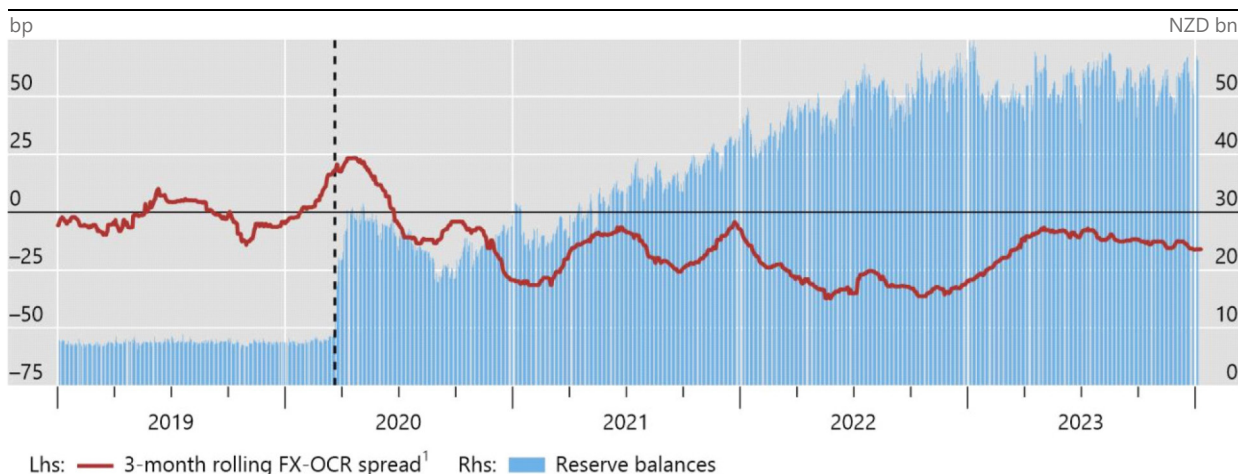
Note: The dashed line indicates 20 March 2020, when the Reserve Bank of New Zealand shifted to a floor system. [1] Spread between the FX-implied one-day interest rate (based on tom-next FX swap quotes) and the official overnight cash rate (OCR); three-month rolling median.

Source: Reserve Bank of New Zealand; Bloomberg; authors' calculations.

To illustrate, consider for simplicity the case in which the participants in the market for overnight funds have access to bank reserves on uniform terms. If so, in an ARS, the overnight rate will settle at the deposit facility rate. But whenever banks hold more reserves than they would like (i.e. they perceive them as costly for some reason),[7] they will push yields on *other* assets down, to an extent that depends on their substitutability with bank reserves.[8] In other words, we are very much in the world envisaged by models à la Tobin and Brainard (1968), where portfolio balance effects rule supreme. The case of New Zealand is instructive here. As shown in Graph 2, when the central bank shifted to a floor system in March 2020 and

---

[7]   In general, these are balance sheet costs, since reserves crowd out other more remunerative assets whenever balance sheets cannot be expanded costlessly (e.g. shadow capital constraints, from business management practices or regulation).

[8]   The 'shadow' overnight rate would be below the deposit facility rate, too, but in principle it would not be observable because the market would break down, as there would be no trading. If the relevant overnight segment also includes participants that do not have access to the deposit facility, the overnight rate will tend to fall below the floor to an extent that reflects the unwanted excess – a 'leaky' floor, as transactions between banks and non-banks or non-banks themselves take place. This is typically the case.

provided abundant reserves to the market, the FX swap-implied New Zealand dollar short-term interest rate was pushed significantly below the overnight interbank rate (the official cash rate). Banks accommodated higher cash holdings by adjusting their demand for other similar instruments.

What about the *transition* from an ARS to an SRS? Our analysis implies that there is a *discontinuity* between the two systems that needs to be taken into account. The nature of the product, as it were, changes as we move from one regime to another – and hence also the underlying determinants of the demand for it.

In particular, it is potentially misleading to make inferences about what the demand for reserves would look like in one regime based on observations from the other. For instance, asking banks the amount of reserves they would like to hold should specify *in which* system. Similarly, whenever the liquidity requirement allows banks to choose its composition – for example, in the Basel III Liquidity Coverage Ratio (LCR) – banks will respond very differently in the two systems. In an ARS, banks will have little incentive to economise on bank reserves when meeting the constraint; the yields on other assets will have adjusted to make those holdings more attractive. By contrast, in an SRS, banks will demand the minimum amount necessary for settlement purposes, in a kind of lexicographic fashion. They will economise on balances as far as possible.

As a result, arguments highlighting the difficulties of operating SRS in the post-GFC environment are probably overstated (Borio 2023). Arguably, the apparent instability in the demand for reserves in that environment, and the corresponding uncertainty, is largely a function of the implementation framework itself, i.e. it is endogenous. This may explain why many central banks around the world have continued to operate an SRS, even after the introduction of the new prudential liquidity requirements (Cap et al. 2020). This includes many central banks in emerging market economies with very large balance sheets, typically reflecting foreign exchange holdings.

Should the authorities choose to move from an ARS to an SRS, they would need to put in place safeguards to ensure that they remain in control of the process and engineer a seamless transition. The key is to overcome the instability linked to the discontinuity in the demand function. They could put in place a narrow corridor via a lending and deposit facility,[9] stipulate in advance the threshold beyond which excess reserves would no longer earn a market rate of return, announce the day of the transition, and absorb the amount of reserves required to move away from the store-of-value region. The corridor could then be widened once the transition has taken place. The alternative of simply shifting back by gently reducing the amount of reserves and hoping to *gradually* and *smoothly* raise the overnight rate may well be unrealistic.

The discontinuity between the two systems given the regime-dependent nature of the demand for reserves is glossed over in standard, in fact ubiquitous, representations of that demand (Graph 3). First, the representation postulates a single demand curve, regardless of the regime. Second, it assumes a well-behaved, exploitable interest rate elasticity. Finally, it ignores the relevance of other yields in the ARS segment, including the effects of the supply of reserves on those yields and their feedback onto the demand of reserves. Rather than moving along a fixed demand curve as supply shrinks, the actual task is more akin to moving across a series of supply-dependent demand curves. The task would be further complicated by the protracted inactivity in the interbank market, which could hinder an obstacle-free redistribution of reserves across banks. Achieving a smooth transition across regimes could be a lot harder than these standard representations suggest.

---

[9] A well-functioning repo or lending facility is critical, especially where the ARS has been in place for a long time, as the discontinuity may be more disruptive. The floor could also take the form of a reverse repo facility.

The stylised standard description of different operating systems   Graph 3

A. Corridor system
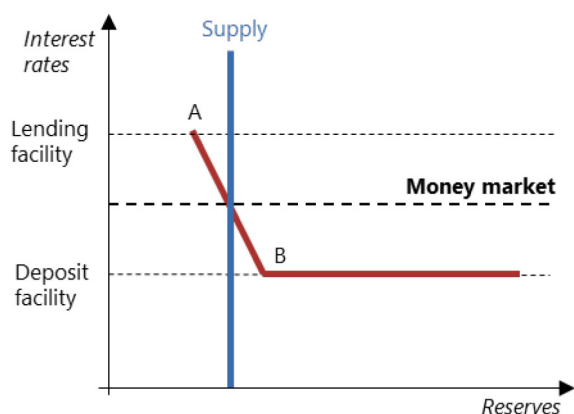
B. Floor system
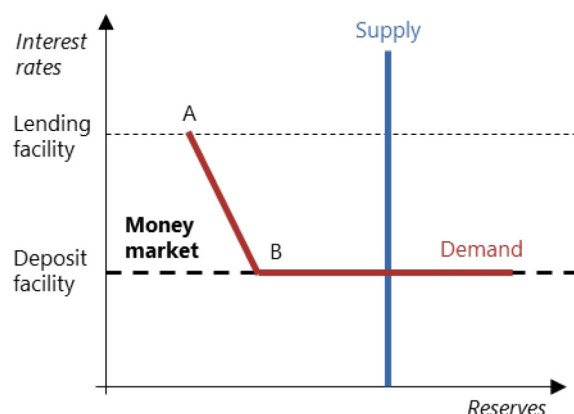
Note: Aggregate demand curves are assumed to be independent of the operating system and the supply of reserves. Point A: if supply falls below this level, banks will access the lending facility which endogenously brings supply back to A. Point B: Point where aggregate demand curve changes from steep to flat. The graph omits the possibility of a 'leaky' floor, an issue encountered in practice with floor systems due to segmentation effects.

Source: BIS.

References

Afonso, G, D Giannone, F La Spada and J Williams (2023), "Scarce, abundant, or ample? A time-varying model of the reserve demand curve", *Federal Reserve Bank of New York Staff Report* No. 1019.

Bech, M and C Monnet (2013), "The impact of unconventional monetary policy on the overnight interbank market", *RBA Annual Conference proceedings*.

Bell, S, M Chui, T Gomes, P Moser-Boehm and A Tejada (2023), "Why are central banks reporting losses? Does it matter?", *BIS Bulletin* No. 68, February.

Bigio, S and Y Sannikov (2021), "A model of credit, money, interest, and prices", *NBER Working Paper* No. 28540.

Bindseil, U (2016), "Evaluating monetary policy operational frameworks", proceedings of the Federal Reserve Bank of Kansas City Jackson Hole Economic Symposium on "Designing resilient monetary policy frameworks for the future", August.

Borio, C (1997), "The implementation of monetary policy in industrial countries: a survey", *BIS Economic Papers* No. 47, August.

Borio, C, P Disyatat and A Zabai (2016), "Helicopter money: The illusion of a free lunch", *VoxEU.org*, 24 May.

Borio, C (2023), "Getting up from the floor", *BIS Working Paper* No. 1100, May.

Borio, C and P Disyatat (2010), "Unconventional monetary policies: an appraisal", *The Manchester School* 78(s1): 53–89, September (also available as *BIS Working Paper* No. 292, November 2009).

Cap, A, M Drehmann and A Schrimpf (2020), "Changes in monetary policy operating procedures over the last decade: insights from a new database", *BIS Quarterly Review*, December: 27–39.

Central Bank of Norway (2021), "Norges Bank's liquidity policy: principles and design", *Norges Bank Papers* No 3.

Greenwood, R, S Hanson and J Stein (2016): "The Federal Reserve's balance sheet as a financial-stability tool", proceedings of the Federal Reserve Bank of Kansas City Jackson Hole Economic Symposium on "Designing resilient monetary policy frameworks for the future", August.

Logan, L (2023), "Ample reserves and the Friedman rule", remarks before the European Central Bank Conference on Money Markets, 10 November.

Lopez-Salido, D and A Vissing-Jorgensen (2023): "Reserve demand, interest rate control, and quantitative tightening", mimeo. www.ecb.europa.eu/pub/conferences/shared/pdf/20231004_mon_pol_conference/Lopez_Salido_paper.pdf.

Maechler, A and T Moser (2022), "Return to positive interest rates: why reserve tiering?", speech at the Money Market Event, Geneva, 17 November.

Markets Committee (2019), Large central bank balance sheets and market functioning, report prepared by a study group chaired by Lorie Logan and Ulrich Bindseil, October.

Reichlin, L, A Turner and M Woodford (2013), "Helicopter money as a policy option", *VoxEU.org*, 20 May.

Tobin, J and W Brainard (1968), "Pitfalls in financial model building", *American Economic Review* 58(2): 99–122.