Liquidity, Settlement Risk, and Systemic Stability

David Marshall
Senior Vice President and Director of Financial Markets Group
Federal Reserve Bank of Chicago

World Federation of Exchanges Annual Meeting
Bangkok, Thailand
September 8, 2017

Liquidity, Settlement Risk, and Systemic Stability

David Marshall[1]

Good afternoon. It's a pleasure to be here. I'd like to thank the WFE for giving me a chance to discuss some thoughts about the role that liquidity plays in markets. I'll particularly focus on the increasing use of liquidity to mitigate settlement risk. But I also want to discuss implications for systemic stability that may be associated with this increasing liquidity dependence.

Please keep in mind that my remarks this afternoon are my own opinions, and do not reflect positions of the Federal Reserve Bank of Chicago or the Federal Reserve System.

I'd like to focus on three points:

- **First**, there are huge benefits to society from expanding the scope of trading in financial markets. But expanding these gains from trade necessarily requires markets to expand their ability to manage counterparty credit risk and other forms of settlement risk.
- **Second**, the infrastructures and processes developed to mitigate counterparty credit risk all do so through ever more intensive utilization of time-critical liquidity.
- **Third**, this increased reliance on liquidity could have adverse implications for future crises. These implications should be recognized and should be addressed if possible.

Let's start with the gains from trade. A foundational insight for the discipline of economics is that the gains from trade can make everyone better off. You may remember Adam Smith's pin factory from Econ 101. Smith's original insight was that trade allows for specialization, and specialization generates massive efficiencies. This same insight is at work in David Ricardo's idea of comparative advantage – that international trade facilitates specialization at the country level, which, again, can generate enormous surplus value. Gains from trade also underlie the importance of well-functioning securities markets, since traded securities allow savings by households to be deployed as capital investment by businesses.

More recently, the growth of derivatives markets allows risk itself to be traded as a commodity. This allows risk to be borne by those most capable of bearing the risk, and having these risk-bearing services priced by market forces.

To give one final example, few would have thought 30 years ago that the right to pollute could be traded as a commodity, like corn or wheat. But the U.S. sulfur dioxide cap-and-trade program, implemented in 1995, achieved the targets for acid rain abatement at a

[1] Senior Vice President and Director of Financial Markets Group, Federal Reserve Bank of Chicago. david.marshall@chi.frb.org

cost estimated at less than 1/5th of the cost that would have been incurred had we used traditional command-and-control regulation. The trading program incented those for whom pollution abatement would be cheapest to bear the lion's share of emissions reduction.[2]

So future growth of societal well-being depends in no small measure on expanding the range of things that trade on organized markets, whether goods, specialized labor, loanable funds, risk, or sulfur dioxide emissions. But expanding these gains from trade also depends on our ability to expand the range of counterparties with whom we trade. In particular, there are major benefits from being able to trade anonymously, or to trade with counterparties whose credit worthiness is less than stellar. But this means that, to fully exploit gains from trade, we need to manage counterparty credit risk. In fact, credit risk is only one form of the more general phenomenon of settlement risk: the risk that, for whatever reason, a counterparty fails to complete its obligations in accord with the agreed upon terms.

So it's no surprise that over the past 40 years we've seen a series of innovations designed to mitigate settlement risk. And what I find fascinating is that they all do so in the same way – through enhanced use of *time-critical liquidity*: liquidity that must be provided at a particular location, in a particular currency, and in a precise time frame measured not in days, but in hours or even minutes.

An important example is the adoption of payment-vs.-payment (PvP) settlement in foreign exchange transactions to mitigate so-called "Herstatt" risk, where asynchronous settlement of the two legs of a foreign exchange transaction allows for the possibility that the later payment leg might fail. PvP, as implemented by CLS Bank, depends on time-critical liquidity provision because participating banks must provide liquidity in the correct currencies during tightly defined funding windows.[3]

To take another example, consider the gradual adoption since the 1970s of real-time gross settlement (RTGS) mechanisms for interbank payments. Prior to the emergence of RTGS, most interbank payments used deferred net settlement approaches. Deferred net settlement, by allowing for netting, is far less liquidity-intensive than RTGS systems. However, deferred net settlement systems cannot provide real-time finality of settlement. This is a major problem with such systems, which was demonstrated by the chaotic conditions following the failure of Bankhaus Herstatt in 1974. At that time,

---

[2] Robert Stavins, Gabriel Chan, Robert Stowe, Richard Sweeney (2012), "The US sulphur dioxide cap and trade programme and lessons for climate policy," August, available online, http://voxeu.org/article/lessons-climate-policy-us-sulphur-dioxide-cap-and-trade-programme.

[3] Bank for International Settlements, Committee on Payment and Settlement Systems, 2003, *Payment and Settlement Systems in Selected Countries*, Basel, Switzerland, April, available online, www.bis.org/publ/cpss53.pdf; and Federal Reserve Bank of New York, Payments Risk Committee, Cross-border Collateral Pool Task Force, 2003, *Managing Payment Liquidity in Global Markets: Risk Issues and Solutions*, report, March, available online, www.newyorkfed.org/prc/files/manage.pdf.

several banks attempted to reverse payments that had been made previously without finality. The result was widespread payments gridlock.[4]

In contrast, a real-time gross settlement mechanism can deliver real-time finality. But its ability to do so is dependent on time-critical liquidity, since an RTGS system transfers funds only if the sender has sufficient liquid balances immediately available.

The G20 swaps clearing mandate also falls into this pattern. A central counterparty (CCP) mitigates counterparty credit risk by concentrating all such risks into the CCP through novation. It then manages these risks largely by requiring adequate initial margin, and by marking positions to market using variation margin cash transfers according to a strict timetable. So reliance on time-critical liquidity for these variation margin transfers is at the heart of CCP risk management.

A similar use of time-critical liquidity to mitigate credit risk is inherent in the G20's movement toward expanded minimum collateral requirements for non-centrally cleared derivatives (again, with daily mark-to-market transfers).[5]

And, looking forward, distributed ledger technology could facilitate near-real-time security settlement– whereby we could move from T+3 or T+2 settlement all the way to T + few seconds. If implemented, this would clearly reduce settlement risk, but at the expense of increasing the time-criticality of liquidity required from both sellers and purchasers.[6]

So we're seeing a pattern that repeats itself over and over. It suggests that there may be a basic principle of financial markets in operation: *once gains from netting are exhausted, further reductions in counterparty credit risk require increased dependence on liquidity provision.*

It's easy to see why this might be so. If you don't fully trust your counterparty, or if you don't have full confidence in a settlement mechanism, then the logical way to control settlement risk is to require that one's counterparty guarantee their performance via provision of collateral, preferably in a highly liquid form — cash, or something that can be turned into cash with high reliability.

---

[4] Herring, Richard J., and Robert E. Litan, 1995, *Financial Regulation in the Global Economy*, Washington, DC: The Brookings Institution.

[5] Basel Committee on Banking Supervision and Board of the International Organization of Securities Commissions, 2013, "Margin requirements for non-centrally cleared derivatives," September, available online, https://encrypted.google.com/search?q=higher+collateral+requirements+otc+derivatives&oq=higher+collateral+requirement+OTC+&gs_l=psy-ab.1.0.33i22i29i30k1l2.11741.18387.0.21401.34.33.0.0.0.0.337.5008.0j17j8j1.26.0....0...1.1.64.psy-ab..8.24.4577...0j0i131k1j0i67k1j0i131i67k1j0i46i67k1j46i67k1j0i10k1j0i22i30k1j33i160k1j33i21k1.xlVeNcY0s6s.

[6] For a formal treatment of these tradeoffs, see: Mariana Khapko and Marius A. Zoican, "Smart Settlement," manuscript, University of Toronto Universitée Paris-Dauphine, February 10, 2017.

The creative uses of liquidity to mitigate settlement risk have brought enormous benefits to society by enabling us to expand gains from trade. But this ever-growing dependence on time-critical liquidity has a downside: it exposes markets to the risk that liquidity may become unavailable. Specifically, market participants may have plenty of illiquid assets, but they may be incapable of converting these assets into the cash required to meet a time-critical deadline. This could happen if, for example, their illiquid assets sell for a price much lower than expected, or if these assets can secure liquidity loans only with a prohibitive haircut.

This sort of liquidity risk is especially problematic in a financial crisis, because crises typically are characterized by pervasive liquidity hoarding. Such episodes are, in effect, generalizations of bank runs or banking panics, although they can occur in markets quite distinct from commercial banking.

Notably, the financial crisis of 2007-2008 was characterized by a series of debilitating liquidity panics that occurred outside the traditional banking system. Rather, panics occurred in the so-called "shadow banking system," a nexus of markets and institutions that performed bank-like intermediation and credit allocation without the costly superstructure of bank regulation. We should keep in mind that the scale of shadow banking activity was huge: by 2007, the size of the shadow banking sector exceeded the traditional banking sector by around 60%.[7] So these shadow banking panics were enormously debilitating for the economy.

Among the more notable instances of such liquidity panics were

- Run on asset backed commercial paper in August 2007
- Run on Bear Stearns March 2008
- Increased haircuts in the tri-party repo market late 2008–2009
- Runs on AIG Sept. 2008
- Run on prime money market funds Sept. 2008
- Run on the commercial paper market Sept. 2008
- Run on the asset backed security market 2008

I don't have time to go through each of these in detail, but they all involved liquidity-dependent entities, sectors, or markets that were unable to obtain sufficient liquidity to continue their normal functioning.

The Bear Stearns run in March 2008 was particularly illustrative. While this example does not specifically focus on the use of liquidity to mitigate settlement risk, it shows how liquidity can dry up in unexpected and counterintuitive ways. Specifically, both Bear Stearns and its regulator, the SEC, estimated that Bear had sufficient collateral to obtain needed liquidity via the repo market and other sources of secured credit. No one

[7] Zoltan Pozsar, Tobias Adrian, Adam Ashcraft, and Hayley Boesky, (2010) "Shadow Banking," Federal Reserve Bank of New York Staff Report No. 458 (*Revised February 2012)*.

anticipated that Bear would be frozen out of the market for secured short-term funding. In the words of Christopher Cox, SEC chair at the time:[8]

> "[SEC liquidity] requirements are designed to ensure that an investment bank holding company can meet all of its cash needs even in the face of a complete cutoff of *unsecured* financing that lasts for a full year. . . . But what neither the [SEC] regulatory approach nor any existing regulatory model has taken into account is the possibility that *secured* funding, *even if it's backed by high-quality collateral* such as U.S. Treasury and agency securities, could become unavailable." [Italics added]

So the increased dependence on liquidity carries with it an increase in liquidity risk. This raises a particularly unpleasant question: does the use of liquidity to mitigate counterparty credit risk actually *reduce* risk in total? Or does it merely *transform* one kind of risk (counterparty credit risk) into a different kind of risk (liquidity risk)? Put another way, have we constructed a system that mitigates garden variety day-in-day-out credit risk in exchange for exacerbated liquidity risk that manifests itself once every 20 years or so in the form of a liquidity panic?

I don't claim to know the answer to this question. But as we double-down on the use of time-critical liquidity to mitigate settlement risk, we clearly need to consider whether, in doing so, we are simultaneously increasing systemic liquidity risk, and, if so, whether there should be a policy response.

A good starting point for this inquiry is to recognize that the problem of scarce liquidity during crises is not new. Liquidity crises are pervasive throughout history. The earliest documented liquidity crisis was a banking panic that swept the Roman Empire in A.D. 33. In the U.S. from 1792 to 1933 liquidity panics occurred every 10 years or so. Furthermore, there is general agreement about how to address the problem of liquidity scarcity: empower a liquidity provider of last resort.

- In the banking crisis of 33 A.D., there was, of course, no notion of a central bank. Instead, the Emperor Tiberius took on the role of liquidity provider of last resort. He deposited the equivalent of $2 billion in surviving banks with instructions for the banks to lend the money out, collateralized by real assets. The crisis ended shortly thereafter.[9]
- In the liquidity crisis of 1792 in the U.S., the nascent central bank of the United States was not yet operative. So Alexander Hamilton, the Secretary of the Treasury at the time, instructed a private institution, the Bank of New York, to buy government scrip on the open market for their own account, with an informal understanding that the U.S. Treasury would provide a backstop if necessary.

[8] Testimony before Congress, April 3, 2008, available online, https://www.sec.gov/news/testimony/2008/ts040308cc.htm.

[9] Tenney Frank (1935), "The Financial Crisis of 33 A. D." *The American Journal of Philology*, Vol. 56, No. 4 (1935), pp. 336-341, and William F. Allen (1887), "The Monetary Crisis in Rome, A.D. 33" *Transactions of the American Philological Association*, Vol. 18, pp. 5-18.

> Again, the presence of this credible source of market liquidity unfroze the financial market, ending the crisis.[10]

The general rules for such a liquidity provider were articulated by Walter Bagehot[11] back in 1873. Bagehot said that to avert panic, the liquidity provider should lend early and freely, to solvent firms, against good collateral, and at a penalty rate. We are comfortable with public sector provision of liquidity, in accord with Bagehot's rules, when the recipients are banking institutions. That's the function of the discount window. It's not that big a stretch to extend such liquidity provision to systemically important, and liquidity intensive, financial market infrastructures, such as payments, clearing, and settlement systems. The key step is to ensure that the institution receiving liquidity be fully solvent, albeit temporarily illiquid. Bagehot makes it clear that under no circumstances should liquidity be provided to an insolvent institution. That principle remains critical to this day.

Specifically, suppose a systemically important infrastructure were subject to a liquidity run. This could happen, for example, if expected liquidity inflows did not materialize. If the liquidity provider of last resort, working closely with the infrastructure's prudential regulator, makes a determination that the infrastructure is solvent, it would be reasonable for the liquidity provider to extend short-term credit. Of course, such credit must be fully secured by high-quality collateral, in accord with Bagehot's rules. And public sector liquidity should only be used as a last resort, after private sources of liquidity have been exhausted. The benefits of such a policy are clear, and the risks to the taxpayer would be minimal. So it is no surprise that this sort of policy is broadly consistent with existing U.S. and E.U. law.

One class of infrastructures where this approach is particularly workable is CCPs. CCPs are relatively transparent institutions. Their risk management assets consist largely of cash, Treasuries, and other low-risk instruments. So it is relatively easy to assess CCP solvency and to determine the quality of collateral. Furthermore, the use of cash margin, which obviously is the most liquid form of collateral, can be further encouraged by allowing customer margin to be deposited in accounts with their central bank, thereby eliminating custodial risk from customer concerns.

The key take-away from this brief discussion of liquidity in financial markets is that liquidity is a two-edged sword. It has proven to be an extremely effective tool to mitigate settlement risk. At the same time, expanded dependence on liquidity may exacerbate systemic risk in a crisis. Like any powerful tool, it must be handled with care.

Of course, there are a number of questions that I have not addressed. What are the trade-offs in extending the range of a liquidity provider of last resort? Are there moral hazard considerations that must be taken into account? What might an optimal

---

[10] Sylla, R., Wright, R. E., and Cowen, D. J., 2009, "Alexander Hamilton, Central Banker: Crisis Management during the U.S. Financial Panic of 1792," *Business History Review*, 83(1), pp. 61-86.
[11] Bagehot, W., 1873, *Lombard Street: A Description of the Money Market*, London: Henry S. King and Co.

regulatory structure look like for liquidity-dependent institutions? Clearly, we need the best thinking on these tough issues. I certainly look forward to your comments and thoughts during the remainder of the conference.