On the Velocity of Bank Reserves

It takes euro bank reserves 27 milliseconds to travel the 4,072 kilometers between Helsinki and Lisbon[1].

The distance between Helsinki and Frankfurt, 2,128 kilometers, takes reserves 14 milliseconds.

Even if arranging an interbank reserve trade takes only 5 seconds, that is 200 times as long as it takes to electronically transfer the funds. And were the ECB a unitary central bank, such as the Bank of England, the reserves would not literally move at all, bank account balances on the single central bank balance sheet would simply be adjusted. Only the message requesting the funds transfer "travels".

Would an increase in the velocity of ECB bank reserves reap large efficiency gains? It would not seem so.

Then what could ECB President Draghi have meant by this?

> "The level of the deposit facility rate can also empower the transmission of the asset purchase program, not least by increasing the velocity of circulation of bank reserves."[2]

Even were reserves to move at the speed of light, the efficiency gains would be trivial.

Mario Draghi studied economics, not electrical engineering, at MIT. So let us delve into economic theory rather than physics in an attempt to glean how increasing the velocity of circulation of bank reserves could enhance the power of large scale asset purchases.

Perhaps the ECB's thinking is rooted in the so called "Quantity Theory of Money" (QTM).

The QTM essentially says that the number of money units in a closed economy (M) multiplied by the number of times each unit is used during a given time period (V), equals the value of economic output sold during that time period expressed in those money units (Y).

Y is usually written as "PQ" where Q is real output measured in a variety of ways (e.g. steel ingots in tons, corn in bushels, services in hours worked) and P is a weighted price index (the exchange rate between the various subcomponents of Q and the monetary unit—say a one ounce silver coin or one electronic reserve balance at the ECB—weighted by the subcomponents' contributions to Y):

$$MV = PQ$$

Clearly, V is not conceptually related to the physical distance traveled per year by the money units but to "turnover". M is easily thought of as an inventory of tokens the economy holds in order to make transactions. We may express the demand for tokens (M) as a proportion, 1/V, of PQ. It is usually argued that V, the turnover rate of the tokens, depends on tastes, habits, and technology which—in turn—are assumed to change relatively slowly. Therefore one may derive a tentative monetary policy conclusion from the QTM. Assuming V is fixed or changes slowly, an exogenous change in the number of tokens (M)

---

[1] A tip of the hat to Andrew Stella for information on the velocity factor of most data cable—between 42 and 72 percent of the in-vacuum speed of light. I've used 50 percent.

[2] *Monetary Policy: Past Present and Future*, speech at Frankfurt European Banking Congress, November 20, 2015.

leads to an offsetting change in Y (measured in tokens). So if the monetary authority can influence M (the *number* of tokens in the economy that are used in transactions) and V does not change materially, then the monetary authorities can influence Y. In the long run, changes in M are thought to influence only P, not Q, but since the ECB at the moment would be happy to have either higher inflation and/or higher real growth, we will not discuss whether money has a "real" or only a price impact, here.

Dr. Draghi's point seems to be that although the ECB has, through its asset purchase program (APP), increased bank reserves (M), little has happened to Y. Thus V must have fallen. Hence wouldn't it be grand were V to rise? And apparently he believes that lowering the ECB deposit rate will do so.

Therefore, Dr. Draghi's claim, viewed through the lens of the QTM is quite simple: "lowering the rate paid on the ECB overnight deposit facility will increase the income velocity of bank reserves—or, alternatively, lower the proportion of nominal income the Eurozone holds in bank reserves".

Since the ECB now determines the nominal quantity of bank reserves—let us say, *fixes* it in nominal terms, the only way for equilibrium in the system to be restored following an increase in velocity is through a change in nominal GDP.

Returning to the basic equation and substituting bank reserves (B) for M, and specifying we are talking about the velocity of bank reserves (V_b) we have $BV_b = Y$. Since the banking system cannot change the quantity of reserves it holds in aggregate[3], an increase in velocity caused by a lower deposit rate will have to result in an increase in Y, voilà, the ECB has achieved its objective.

It all seems so simple.

And indeed, the basic QTM, about which Keynes said "The truth of this, properly explained and qualified, it is foolish to deny"[4] is quite simple. But perhaps too simple.

As Keynes implied, the assumptions that are necessary for the QTM to hold are extremely strong (and therefore are very important to grasp and explain). As a corollary, without relaxing the assumptions, it is difficult to make the theory empirically operational in an advanced financial system. In particular, many different types of messages/tokens/securities and implicit contracts are used to make not only retail payments but also, and to a much greater extent, wholesale payments.

The point about wholesale interbank payments is very important. In the advanced economies, gross payments using bank reserves are roughly 50 times GDP meaning the vast proportion of those payments are for reasons other than paying for value added—e.g. to settle securities transfers.

For example, in the US, on an average day about US$ 3.5 trillion in payments are originated over Fedwire amounting in 2014 to US$ 885 trillion compared with GDP of US$ 17 trillion. So if we are thinking of the QTM as a demand for money function, it is rather obvious that demand is more likely to

[3] Let us assume for the moment that we are at a corner solution where demand for physical banknotes is zero and the central bank does not engage in foreign exchange intervention so B is fixed.
[4] See chapter 2, *A Tract on Monetary Reform*, John Maynard Keynes (1924).

be related to the value of *transactions* rather than to economic value added. Indeed, if we compare the value of US bank reserves and GDP over long periods of time it is evident there is little to no relation (see below). Nor is there a statistical relationship between bank loans or credit and bank reserves.

But that does not stop us from writing out equations that, by definition, must be true.

In other words, coming back to my opening remarks, velocity is not something we measure. It is a definition. Velocity cannot be observed independently of B and Y, both of which can be measured independent of our knowledge of velocity. Thus any remark we make about velocity is logically identical to some remark we make about B or M and Y. The statement "monetary policy has not worked because velocity has fallen" is identical to the statement "monetary policy has not worked because there has been no impact of a change in money on nominal GDP". Neither of those are explanations of anything. They are merely descriptions of the change in components of a theoretical identity.

We do not arrive late at a party and say "my automobile is responsible because its average velocity fell", nor "my automobile moved less rapidly than last time I pressed that much on the accelerator". Both of those statements may be true. But as explanations and diagnostic guides they are vacuous. Perhaps there was a storm, a traffic jam, the driver spent a good deal of time in neutral gear without realizing it, or the engine malfunctioned. All of those represent the beginning of an explanation.

Let us move now into the empirical plane. There is one extremely important conceptual error that has been passed down through generations of students of monetary economics. That is, quite simply, that central banks conduct monetary policy by setting the money supply. I would conjecture that not a single advanced country central bank with an independent monetary policy operated this way between 1980 and 2005. It was taught that: the central bank changes the money supply and the impact on nominal GDP is the multiplicative product of the change in money and the velocity of money. In other words, money was a causal factor. Wrong. The causality ran the other way. Changes in nominal income (and securities markets transactions)[5] determined changes in money.

Central bank money consists of banknotes and bank reserves. During my 25 years at the IMF, apart from countries in the former Soviet Union and a few other quite LDCs, no central bank tried to control the quantity of banknotes in circulation. Banknotes are supplied on demand to the public. So any central bank aiming at a monetary base target always aimed to influence the bank reserve component of the monetary base (approximately 5-10 percent of base money in general).

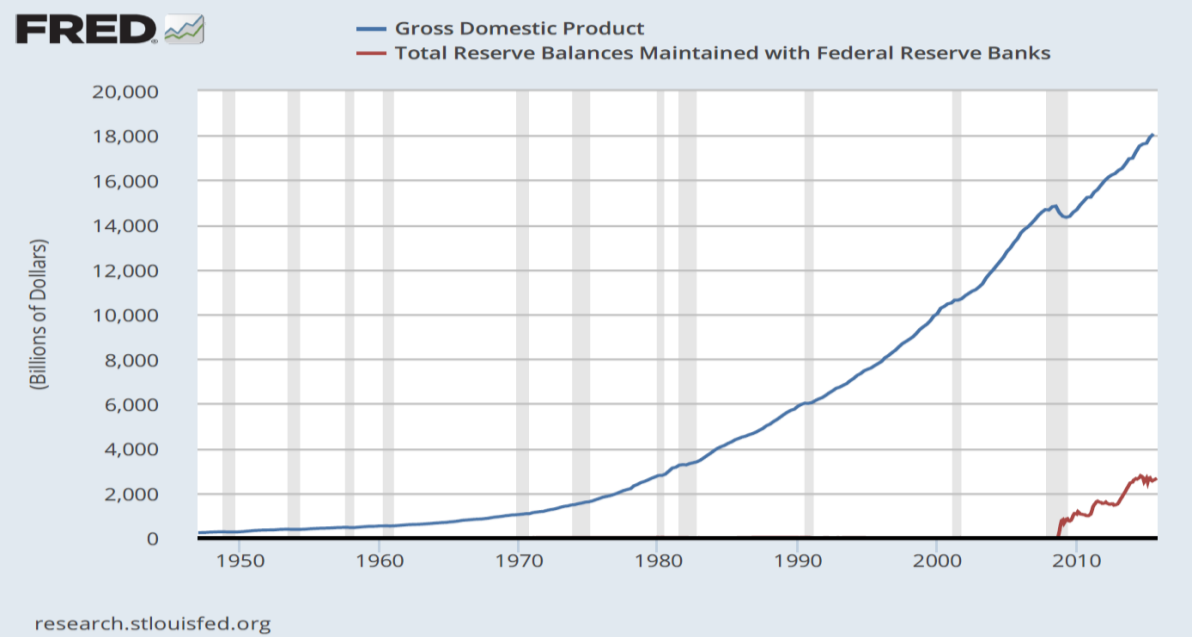
No advanced country central bank tried to control the quantity of bank reserves during the 25 years leading up to the Great Recession[6]. They set, as a target, an overnight interbank interest rate. In order to do this, they allowed the banks to determine the amount of reserves in the system. As is the case with any monopolist, central banks had to choose between setting the quantity supplied or the price at which

[5] Back in the 1960s there was a significant debate over whether the quantity theory should be applied to income or total gross transactions. Although cash is generally used for final purchases, bank reserves support wholesale payments—primarily related to securities market transactions and intermediate purchases. Indeed, in the US, the long term correlation between GDP and bank reserves is negative.
[6] Ulrich Bindseil's book *Monetary Policy Implementation: theory, past, and present* is the classic here.

it supplied reserves—picking a point on the demand curve for reserves. All central banks chose to set the price, economic theory, by the 1970s, demonstrating that rationing by price is generally more efficient than rationing by quantity.

In sum, the quantity of the monetary base, B, was determined by the market, and supplied by the central bank at its chosen interest rate. The interest rate influenced Y, and Y in conjunction with the interest rate and payments systems technology, determined M and maybe B. So when we look at simple movements of Y and M and B, we need to understand that Y is the dog and M or B is the tail. B—the variable directly controllable by the central bank—is not the tail that wags the dog. Indeed, as stated above, the US dog simply does not wag its tail…there is no relation between bank reserves and GDP[7].

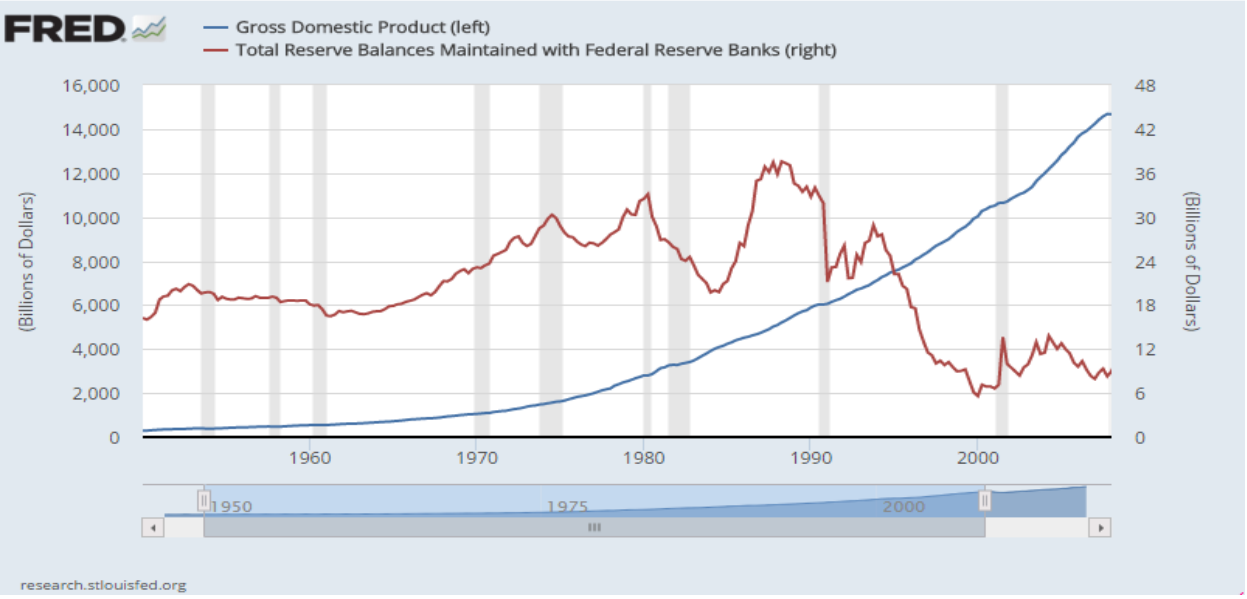
Let us take a look at how B and Y have evolved during the last 65 years in the United States.



Total reserves fell from $ 16.5 billion in January 1950 to $ 8.7 billion in January 2008 and are not visible in the chart until the post-Lehman surge in September 2008. The annual "velocity" of bank reserves rose from 17.3 to 1,636.2 during those same 58 years. Alternatively, the proportion of nominal income held in bank reserves declined from 5.8 percent in 1950 to .06 percent in 2008.

Using a different scale for GDP and Total Reserve Balances, we have the picture for 1950-2008:

[7] There is between GDP and demand for currency and hence between "central bank money" and GDP. But the relationship runs from GDP to currency, not the other way around, for the reasons given.



It certainly does not seem that there is any correlation between the two series. In other words, the process driving US GDP and the process driving demand for B seem independent[8]. It is important to note here that at least for the two decades prior to the Great Recession, central banks provided banks an unrestricted quantity of reserves—at or around the overnight policy rate. So the quantity of B was determined by the market directly. For those same decades, all advanced country central banks held the market "short" reserves, meaning that banks had to routinely borrow reserves from the central bank to meet their inventory demand for reserves. In this way, the policy rate—tied to the overnight central bank lending rate—was thought to "bite" the market which is another way of saying that current and expected future policy rates were thought to be the marginal cost of bank funding and the fulcrum of the monetary policy (more appropriately thought of as interest rate policy) process.

Prior to the crisis, the only actively used policy instruments were short term liquidity supplying repos. The Fed, for example, like its advanced country peers, kept the market "short" reserves and influenced the overnight rate via lending operations. In the aggregate, banks were compelled to rely on small amounts of credit from the central bank to obtain their desired reserve balances. This facilitated the attainment of central bank interest rate targets even with small operations.

As of December 5, 2007, the Fed was providing $47 billion in 7-day and 14-day repos to enable the market to meet its demand for $16 billion in overnight deposits (see Table 1 from Stella (2015) below). The then current FOMC target for the fed funds rate was 450 bps and the daily effective fed funds rate on December 6, 2007 was 449 bps.

[8] Elsewhere I've looked at the correlation between lagged changes in bank reserves and bank credit in Japan and the US. In both countries the correlation coefficient is zero for several decades before the Great Recession. Including more recent data the correlation turns negative simply owing to the fact that reserves have expanded enormously and credit has declined.

Table 1: Consolidated Balance Sheet of the Federal Reserve Banks

December 5, 2007

(in US$ billions)

Assets		Liabilities	
US Treasuries	780	FR Notes Outstanding	782
Liquidity Providing Repos	47	Bank Deposits (Overnight)	16
Net Other Assets	50	Reverse Repos	37
		US Treasury Deposits	5
		Equity	37
Total Assets	876	Total Liabilities	876

Source: (Stella(2015), *Exiting Well*)

The quantity of bank reserves demanded (and passively supplied) in the US may seem shockingly low. Conversely, the income velocity of bank reserves may seem shockingly high at 1,000+. Indeed, back in the 1980s, the income velocity of M1 was thought stable at about 7. Reconciling the small quantity of reserves within the monetary base and broader monetary aggregates is possible with a correct understanding of the role bank reserves play in modern financial systems.

Bank reserves are used *only* in the interbank (wholesale) market to facilitate intraday payments.

Modern financial systems are able to process trillions of dollars, euros and sterling payments *every day* with a very insignificant inventory of reserves. For example, transfers of bank reserves executed through Fedwire, the primary US real time gross settlement system (RTGS), amounted to about $900 trillion during 2015 compared with US GDP of about $18 trillion. Thus annual gross bank reserve transfers through Fedwire are approximately 50 times what would be needed to “buy” US GDP. Obviously, quite a large segment of Fedwire transfers are to facilitate securities trading; and the marginal cost to use Fedwire to facilitate settlement is close to zero—and is unrelated to the size of reserve balances.

Select Fedwire Payments Statistics (in US Dollars where applicable)

Time Period	Average Daily Transfer Value (in billions)	Average Daily Number of Transfers	Average Daily Value per Transfer (in millions)	Bank Reserves (in billions)	Average Daily Turnover of Bank Reserves
9/2015	3,433	580,261	5.92	2,734	1.2
9/2007	2,887	540,143	5.34	32	90.2
1987	605	207,839	2.91	36	16.6

Sources: Fedwire Funds Service; FRB release H.4.1; and author's calculations[9].

Therefore, if we wished to have an idea of the "transaction velocity" of bank reserves pre-crisis it was 90 times *per day* or approximately 23,000+ times per year. Alternatively, banks held overnight about 2 percent of the average daily amount they sent through Fedwire in September 2007.

The average daily turnover of bank reserves in the US is now 1.

Is it really a puzzle as to why daily "velocity" collapsed from 90 to 1? No, the answer is quite simple. The Fed bought $3 trillion in securities from nonbanks and paid for them through Fedwire by crediting the nonbank sellers' clearing bank with reserves. JPMorgan did not *choose* to increase its reserves from the $2 billion it held in 2006 to $447 billion at end-March 2015, it simply resulted as an unintended consequence of Fed purchases of securities from its customers[10].

The absolutely fundamental point to understand about bank reserves now in the US, the Eurozone and everywhere that large scale asset purchases have taken place, is that their quantity is no longer determined by bank demand. The quantity of bank reserves is determined by a completely different process—the net magnitude of LSAP. Any theoretical construct that was useful to understand the bank reserve market during the last two or three decades needs to be completely rethought. For twenty years, the quantity of bank reserves was determined by RTGS technical requirements and procedures, the volatility of intraday bank inbound and outbound transfers, the creditworthiness of counterparts, the functioning of the interbank loan market and its micro structure, and the cost of acquiring reserves from central banks. Now they are determined as the mirror image of large scale asset purchases (LSAP). A consequence of this is that commercial banks are now huge *net lenders of reserves to central banks*.

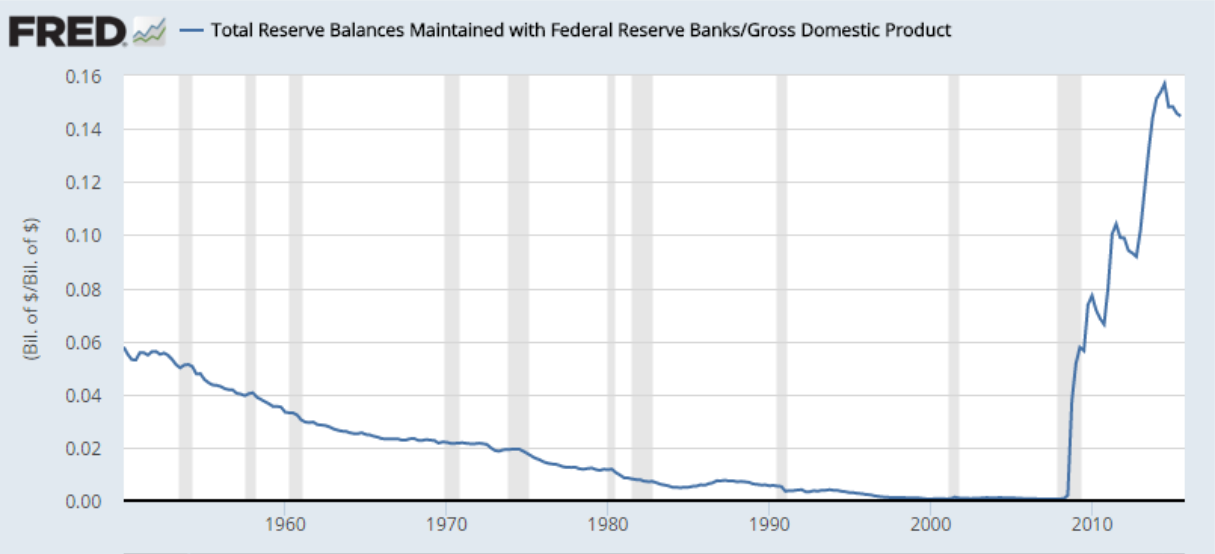
Central banks having engaged in LSAP have gone from price setting monopolists in bank reserves (net lenders) to price and quantity setting monopsonists (net borrowers) in bank reserves. How can a monopsonist set both price and quantity independently? Only if the market is not clearing. In this case, the demand for bank reserves at the current price is much less than the quantity set by the central bank.

The US commercial banking system is financing the Fed's inventory of securities involuntarily. The same process is underway in Europe—a bit further behind. There is nowhere else banks can place reserves—by law and definition—than at a Federal Reserve Bank or at a Eurozone member central bank. So

[9] Data on reserves is for September 9, 2015 and for September 12, 2007.
[10] For more on this, see Stella (2015), *Exiting Well*, on stellarconsultllc.com/blog.

whether the Fed deposit rate (the rate at which it pays to "borrow" from banks to finance its assets) is 25 bps or 125 bps makes absolutely no difference to the aggregate amount of reserves outstanding. The marginal cost of bank financing is no longer the policy rate, it is the rate a bank must pay on its marginal liability. The same holds true of the ECB. Whether it pays 10 bps or -50bps, it will not change the aggregate amount of reserves in the system (other than by attrition—reserves will fall as they are "taxed" away just as reserves would rise with positive interest paid all else constant).

As an unintended consequence of Fed LSAP, US banks now hold trillions of dollars of unwanted reserves—and the proportion of national income "held" as bank reserves, as of July 2015, reached 14.5 percent, a post-World War II record. The US ratio, B/Y, is shown below.



The annual "income velocity" of bank reserves has fallen to 6.9.

What is actually the clearest aspect of LSAP and the ECB's APP? Bank reserves will rise and remain large unless and until the assets acquired are sold or mature[11]. As a matter of balance sheet arithmetic, this means either banks must decrease other assets or increase liabilities (funding). In other words, in the optimistic scenario where the banking system balance sheet expands its asset side euro for euro with the APP (other lending remains the same), banks must expand their financing euro for euro with the APP. Everything else constant, this means banks must raise deposits and equity. So interest rates on deposits and the return on bank equity would have to rise as banks' funding demand increases. But we do not observe that happening. Indeed, the way LSAPs work has to be through an adjustment in the nonbanking portion of the financial system. The asset purchases work to lower market rates and drive up asset prices to the point where the risk/reward tradeoff establishes an equilibrium between new (lower) rates on financing provided to banks and the new (lower) rates of return on alternative assets.

[11] Central banks can (and have) offered other liabilities such as term deposits which are not overnight liabilities and therefore not "reserves". As long as these liabilities are to banks, bank balance sheets will remain large.

Banks are essentially locked into holding a quantity of reserves and cannot adjust the yield on those assets (loans to the central bank) and locked into an equivalent amount of financing for those loans with the rate determined by the market. The nonbank market can adjust its portfolio freely with the constraint that it must finance the banking system with increased deposits. *Its* action reestablishes an equilibrium set of interest rates and yields. Banks are essentially passive bystanders.

In other words, LSAP presumes a market for asset class A. The central bank buys a quantity of A at a market clearing auction (in exchange for B). The quantity of A in the market falls. In order to equilibrate demand and supply in that market, the price of A rises and the rate of return on A must fall. This leads to a cascading impact on other asset classes as their prices are bid up and rates of return fall to equilibrate the system. Among those assets are nonbank loans to banks or, if we prefer to use the deposit nomenclature, the price of deposits rises and interest rates on deposits fall. But arbitrage stops there. Banks "lend" the central bank all their reserves at a fixed price, the deposit rate.

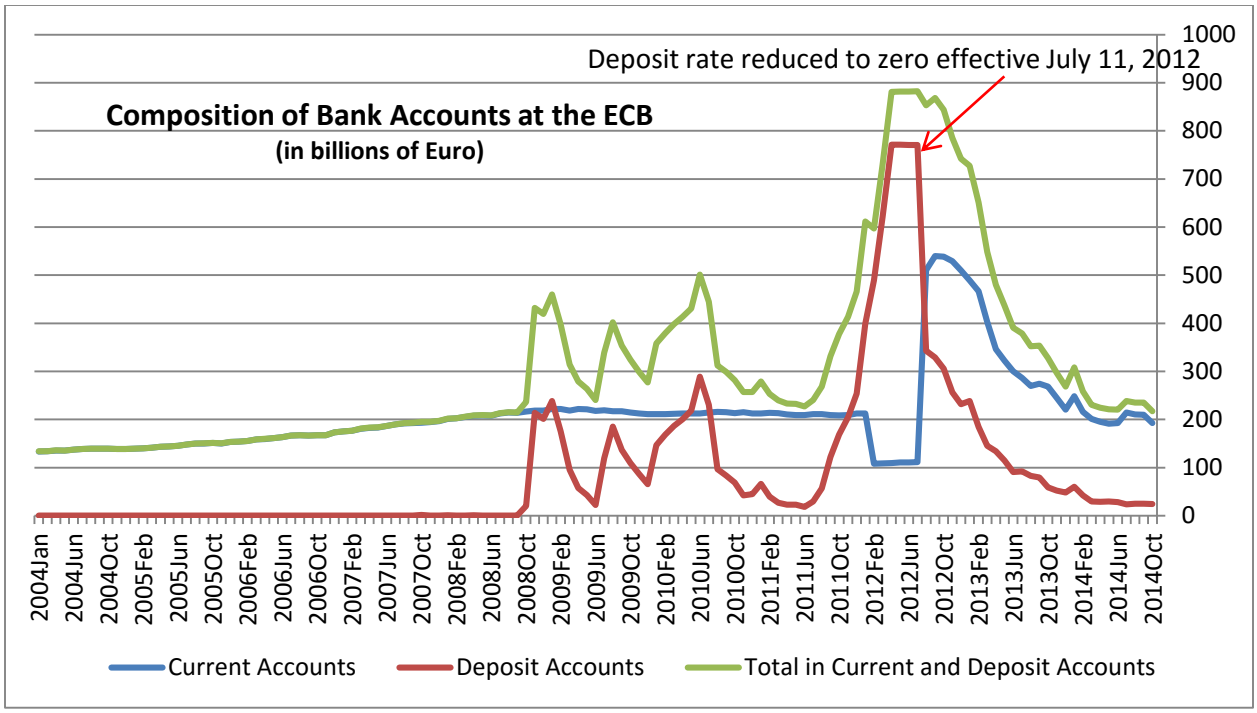
All this implies that looking for clues to the effectiveness of negative deposit rates in the behavior of bank reserves is mistaken. The "market" for reserves is absolutely no longer a "market".

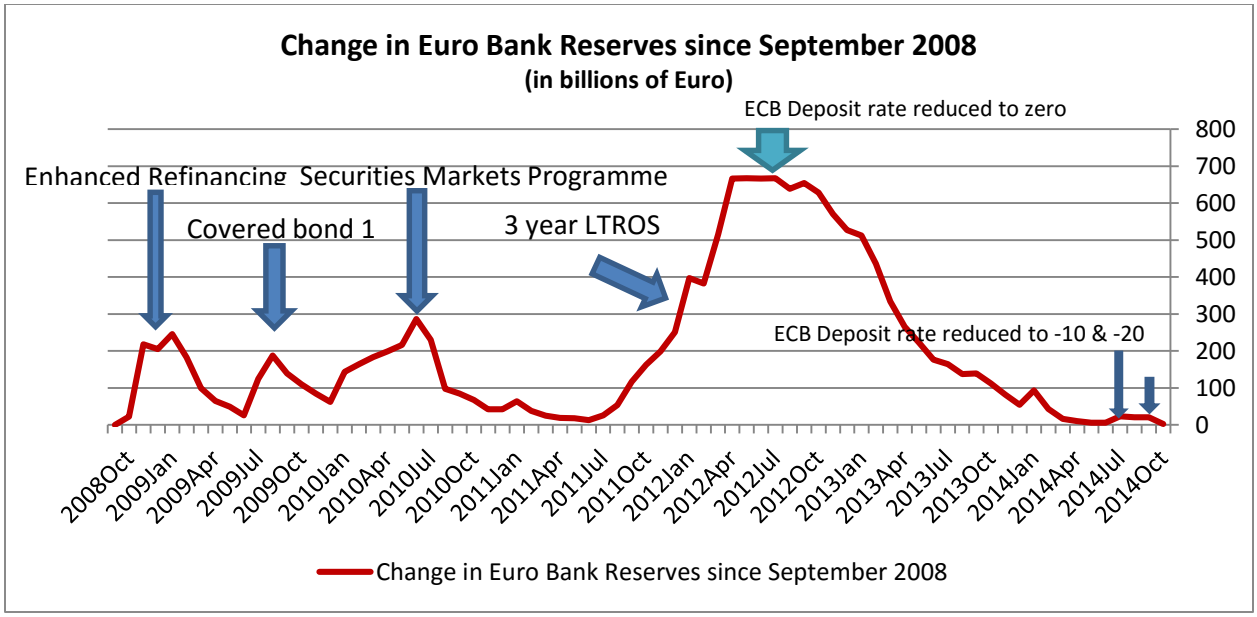
Let us return then to the equation/definition that must be true, with some rearrangement:

$$B = (1/V_b)PQ$$

What happens with LSAP is that the supply of B becomes exogenously determined by the central bank and the central bank expands B well beyond demand at a zero nominal lending rate. So B >> than β, where β is defined as the demand for bank reserves at a zero nominal lending rate. Thus we will observe B, and might conclude erroneously that the demand for bank reserves has increased, that is the desired proportion of income held in bank reserves has increased as the result of a market equilibrium process. It has not. In fact, since the banking system is lending to the central bank now, rather than borrowing, the demand for bank reserves at negative rates is even less than that at zero.

Indeed, when banks were still able to determine the supply of reserves in the Eurozone, once deposit rates were cut to zero, banks immediately began reducing their LTRO borrowing, thereby reducing their reserve holdings. The demand for reserves fell—essentially owing to Draghi's "whatever it takes" and the reduction in deposit rates and availability of the LTRO prepay option. Observed reserves plummeted.

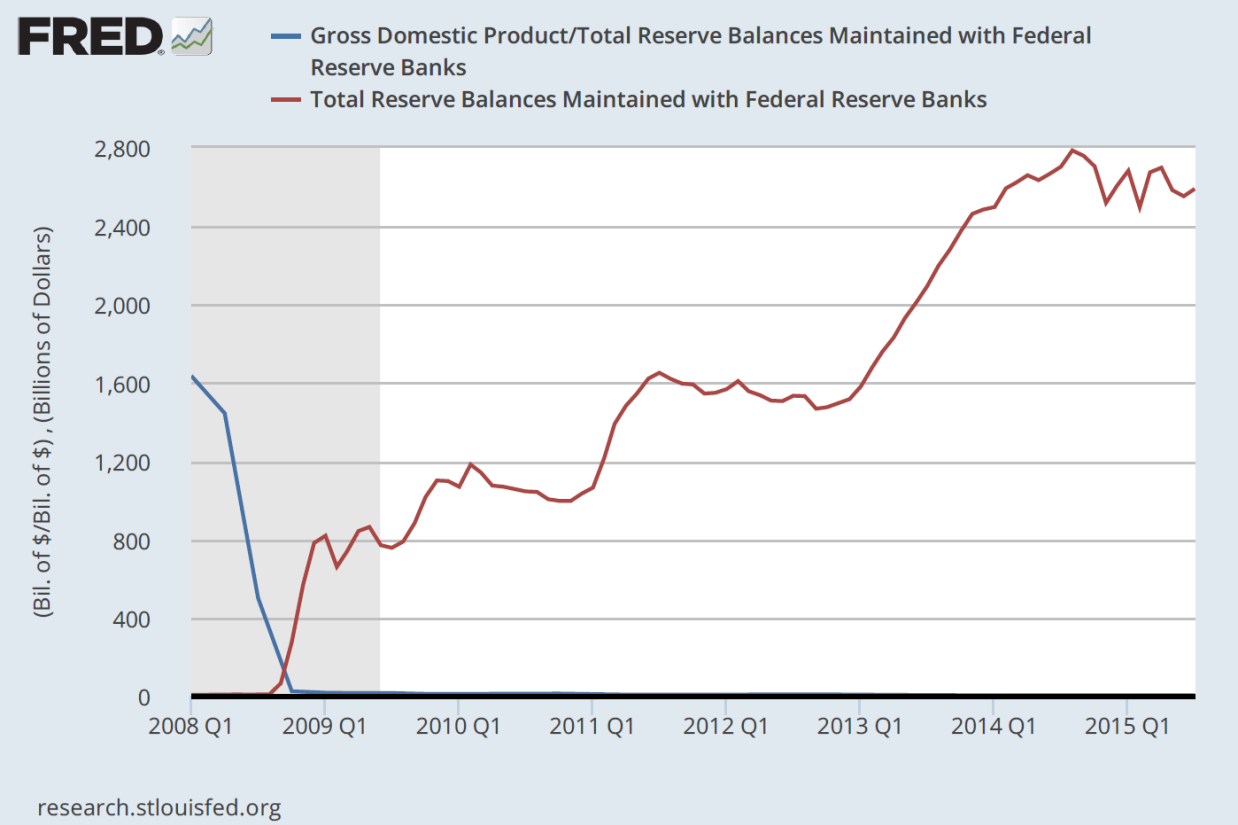




Source: *What the ECB has Done*, stellarconsultllc.com/blog.

The reduction of the deposit rate below zero reduced reserves back to pre-crisis levels.

The behavior of bank reserve velocity post Lehman is shown in the US data below. Bank reserves rose well beyond what was necessary for smooth functioning of the payments system and bank reserve "velocity" collapsed from over 1,600 to 6.



What these several last charts show is that the ECB went through a period of lowering the deposit rate when banks were still able to control the quantity of reserves demanded. They reduced their demand for reserves as the deposit rate was cut (reserve "velocity" increased). They are now entering into a period where all banks with access to alternative funding have repaid their loans to the ECB. The start of the APP has increased their reserves involuntarily at the ECB and we are witnessing a collapse in bank reserve "velocity" as was seen post Lehman in the US. The ECB would like to increase the velocity of circulation of reserves while the market for reserves is out of equilibrium.

The market for bank reserves is out of equilibrium, commercial banks have transitioned from small borrowers from central banks to enormous lenders to central banks and some—the ECB, and BOJ have decided they will pay only "negative" rates to their captive lenders. While negative deposit rates may "work" through certain channels, it is important to understand that combined with LSAPs they will not work via a direct increase in bank lending to the real economy.

Peter Stella
http://stellarconsultllc.com/
4 March 2016