

If you had to write a *brief* note on the Quantity Theory, what would you say? v 1.1

James Forder

3 March 2023/2 Nov 2024

Consider the Quantity Theory *equation*, or ‘the equation of exchange’ as it is often called:

$$MV = PT$$

Where M is the quantity of money, V , the velocity of circulation, P is the price index, and T the number of transactions. Definitional issues of some sort arise in the case of each variable. To relate the equation to any actual economy, some specific definition of money is required. The measure known as ‘M2’ is often used, and it is routinely, and well enough though slightly loosely described as ‘currency plus bank deposits’. Intuitively speaking, ‘velocity’ means something like ‘the number of times each piece of money is used in a time period’. That is all very well when all money takes a physical form, but does not quite make sense for bank deposits. Never mind for now. The price index means the index of prices of the things the sale of which count as transactions. Usually, ‘transactions’ means ‘final transactions’. That is, if you like, the transactions of goods and services that count in the calculation of national income. (So, for example, the value of intermediate or semi-finished goods is not included in national income as they would then be double-counted when the finished version was also included. So here, their sale is also not a ‘transaction’).

The two sides of the equation then give alternative ways of calculating the total value of transactions (and actually the equation is an identity). One could perfectly well use different definitions of money and transactions so long as appropriate values were given to velocity and the price index, and everything would work out. In the case of money, it is a matter of dispute which is the most useful definition. In the case of transactions, I doubt anything but ‘final transactions’ will lead to much enlightenment, so I stick to that and the associated ‘income velocity’, as it is sometimes known.

The Quantity *Theory*, as it is inherited from Classical economics, and revived (or revived with additions) by Friedman (1956) is the proposition that V and T are more or less constant, easily predictable, or exogenous, and that changes in M determine changes in P . Thus the Quantity Theory is often seen principally as a theory of inflation, and tends to lead to its being blamed on the government (or central bank) on the basis that it controls (or can if it tries) the quantity of money. (It would of course be admitted that such things as the 16th century shipping of gigantic quantities of silver from South America raised the price level in Europe, without having anything to do with the government ‘printing money’). In extreme versions, rather precise timing is alleged, so that it will be claimed for example, that an increase in M will raise P eighteen months later. In looser versions, it is accepted that such effects are subject to what Friedman famously called ‘long and variable lags’, but still they are supposed to show up at some point. Along these lines, the proposition that the relationship between M and P is stable and predictable enough to be the basis of policy, and in particular anti-inflationary policy became known as ‘monetarism’ in British policy debate in the 1970s and was (sort of) adopted by the Thatcherites.¹

¹ I am not at all sure that is a good representation of Thatcherism, though it is completely standard. I would tell you more, but I have to be brief.

That is one view of it. As Friedman also sometimes said of the Quantity equation, it offers a framework for analysis. It points a way to understanding what is going on by thinking in terms of these four variables and their relationship. Looking at it like that, anyone should be ready to accept that if one is telling a macroeconomic story about any of the four variables, one ought be ready to fill in the story of what is going on with the other three. In telling these stories, one faces plenty of possibilities as to the determination of the variables, and matters of causation. So, for example, T will vary over long periods because of economic growth. That calls for a small arithmetic adjustment, but nothing difficult. T will also vary in the business cycle, being lower when unemployment is higher. So, there is something of a mechanism for unemployment to *cause* inflation (constant M , constant V , falling T . What's left?) That particular one is probably not really worth worrying about as the changes (or volatility in uncertain conditions) in both M and V will be greater than those in T . If it is not only cyclical unemployment that we consider, but there is the possibility of underemployment equilibrium of the kind described by Keynes, then of course, an increase in M might perfectly well increase T rather than P .

Another point is that if we choose to treat V and T as more or less constant, it is natural to think of changes in M as causing changes in P . An easy story along monetarist lines would be that if there is an increase in M , the private sector will temporarily be holding greater money balances than it wants.² It might respond to this by spending more, and this additional spending will raise the price level, thereby reducing the real value of the money stock. Equilibrium is achieved when the price level has risen to the point that the private sector wishes to hold the quantity of nominal money balances that equates to the original level of real money balances (at which point the extra spending stops).

On the other hand, a correlation between M and P might be explained by causation going the other way. Money consists of currency plus bank deposits, and deposits are created by banks (not just the central bank) when they lend to customers. But banks lend nominal amounts, and the volume of lending that is appropriate will therefore depend on the price level. For example, it is good credit- policy to lend more to a borrower with a large *nominal* income (or turnover) than to one with a smaller one. So if all incomes rise in nominal terms, other things equal, lending will rise and P has causal effect on M . (The policymaker can probably put the breaks on, but the *first* question is how things go when the policymaker sits still). Similarly T (or planned transactions) might affect M . As Kaldor (1970), thinking 1970s-style of currency rather than deposits as the crucial kind of money, asked – is it to be believed that the increase in currency in the hands of the public every December is the *cause* of the Christmas spending spree? The possibility of 'reverse causation' is widely accepted (including explicitly by Friedman) but the monetarists tend to diminish its importance, whereas the post- Keynesians often treat it as a key part of their outlook. The possibility of exogenous determination of P arises from the old idea of 'cost-push inflation'. In a modern form it appears in such analysis as the Carlin and Soskice (2015) 'WS/PS' framework – there is no money in that.

And then there is the question of the determination of V . In practical terms, if one wants to know the value of V (in an economy with an advanced banking system), one finds out by dividing PT by M . That no doubt makes it seem that V is simply a residual determined by the

² In equilibrium, all the money that exists must be willingly held by someone. Therefore if desired money balances are not the same as the money supply, there must be disequilibrium, and *something* will have to change.

operation of the rest of the system. But that can hardly be the whole story – V must have economic determinants.

Most of the time, the trick to thinking about those determinants is to see V as being like an inverse of the demand for money. The demand for money of an individual (household, business, whatever) measures the value of money (i.e. 'currency plus deposits' or whatever definition is being used) that the individual wishes to hold – it is the 'demand for money holdings', or '... for money balances', if you like. It is the outcome of an asset-allocation decision: An individual of a certain wealth, income, and behaviour pattern, wishes to hold currency plus deposits to a certain value, equity to some value, real estate to some value, etc.

A natural starting point for a theory of the demand for money is that the principal reason for holding money is that it is needed for making transactions (it provides 'transactions services'). In the same way the reason for holding real assets is that they provide services – real estate provides housing services, a washing machine provides washing services, etc. And a primary reason for holding non-monetary financial assets is that they offer financial returns in the form of interest and capital appreciation (or greater such returns than deposits do).

To simplify, we tend to restrict attention to the allocation between money and other financial assets. Then it is natural to suppose that two key variables determining V will be an individual's expenditure (per unit of time), and the rate of return on other financial assets. Expenditure is naturally linked to income – so those who have a large income are expected to have large expenditure each month, say, and consequently to have a high average bank balance during the month. The point that a high rate of return ('interest rate' in the usual shorthand) on other financial assets will naturally encourage everyone to economize on money balances. Sticking with just those two, it is probably natural to think that one could estimate a 'money demand equation' giving desired money balances in terms of income and the interest rate.

Now consider the demand for money per unit of income. A household (say) has a desired money balance per unit of its income. Its income and expenditures are flows into and out of that balance. There are alternative arrangements that will finance the same total value of expenditures. Consider a household that is paid once a month and pays all its bills once a month, and otherwise maintains no money balance. If it pays the bills the day after receiving income, its average balance is about zero. If it pays the bills 29 days after receiving income, its average balance is approximately equal to its income. So there is a vast difference in its demand for money. Now consider the case where every economic agent behaves in the same way. In the first case, it is money-in, money-out for everyone; demand for money is very low, but velocity is very high. In the second, everyone sits on the money for a month and it circulates very slowly, but demand for money is high.

Hence, the velocity of circulation reflects (inversely) the demand for money balances. This means that most of the time, we can understand the determination of V by theorizing about the demand for money per unit of income. Up to now, the only variable is the rate of return on other financial assets, but that makes the point. If the interest rate rises, velocity is expected to rise. The reason is that at high interest rates, individuals are incentivized to economize on money balances and to do that they seek to make acquisitions of money (by

income payments, or sales of other assets) simultaneous with expenditures. Then, their average money balance is low; the same is true of everyone; and consequently money circulates fast.

Clearly, then, changes in payments technology, or the liquidity of the stock market, etc will affect money demand and hence velocity, though mostly only slowly. It could also be that the 'liquidity services' of money have non-unitary income elasticity. So, the demand for money might rise as income rises since liquidity services are a 'luxury good', for example. No doubt payments technology improves as part of the growth process, so in those two ideas we have one pushing each way in the long-term trends. (In the American data, broadly speaking, velocity declined up to the second world war and rose afterwards).

So far, so good, but what else might affect the demand for money? Friedman and the monetarists took the view that there was not much, and almost nothing of great interest. A few things can be admitted, such as the cyclical nature of velocity. Payments tend to speed up in booms, perhaps because the optimism of such times leads to free-spending. In downturns payments tend to be delayed, so velocity slows. But things like that are small matters, easily accommodated into a sensible view.

The postwar British Keynesians went much farther down the road of saying that V was more or less unpredictable. It could almost be said that it really was just the residual. T was determined by employment, and hence by the vagaries of private sector saving and investment decisions and employment policy; P was determined by costs (with cost-push inflation a key problem); M was unimportant, because V would adjust as required.³ There was a whole collection of arguments about this – I suppose one would look in the report of the Report of the Radcliffe Committee (1959), from which it became something near to official doctrine. Here are two thoughts. One is fairly clear in Keynes (1936). That is that demand for money will rise when financial investments appear unattractive. If one also takes the view that attitudes to investments are themselves volatile and unpredictable, the demand for money and hence velocity will also be unpredictable. A second argument characteristic of the times is that there are readily available substitutes for money, so any attempt to restrict M will result in the use of these substitutes, and an increase in velocity. A likely candidate is trade credit. Consider the situation where the central bank succeeds in reducing the quantity of money by causing the banks to limit loans. Businesses are thereby deprived of money. But are they deprived of the ability to do business? 'Trade credit' is the credit one business gives another, for example, by delivering goods in advance of payment. An expansion of trade credit might create a substitute for money. A business might be operating, just as before, on credit, but it would not be credit from the bank. One could go further. In sufficiently extreme cases, businesses might issue scrip in the form of paper claims on their own output. So long as one continues to count all the transactions (as is natural for a proper understanding of GDP) then one is forced to the view that V is increased. It is not actually that the money is moving faster, but that some of its job is being done by other means. Still, the measured V , being PT/M will rise. This sort of reasoning leads to the conclusion that V is unpredictable, depending on much wider considerations of

³ Notice that this is also what goes on in Carlin and Soskice (2015) – T is naturally understood as being determined by the level of employment, and P comes out of the WS/PS framework. Neither of those is monetary. Something must be determining M , though we do not know what, and V is then presumably doing its duty by making sure the Quantity Equation holds.

liquidity and how purchases might be financed than the in the simpler picture. And consequently, as the British Keynesians thought, there is not much point in worrying about M (or certainly not in trying to use it as a policy tool).

That leaves us with what one should think about the demand for money, or the stability of velocity. It is clear that the 1970s monetarists were far too optimistic about getting good predictions of the rate of inflation over short time periods. The demand for money is affected by the business cycle so a recession like that of 1981 was a hopeless time for forensic monetarism. On the other hand, the old Keynesian idea that velocity was simply not worth thinking about or studying seems to be a very extreme view. The thought that broadly, there is stability in behaviour and we should expect money balances to reflect income more than anything, with allowance for interest rates and slow-moving considerations such as payments technology, seems very reasonable.

So, one can consider various issues. One might be the Carlin-Soskice deflationary spiral. Apparently P and T fall faster and faster with nothing to stop them. M may be falling as well, but will it fall without limit? If not, does this mean that Carlin and Soskice think that V will fall without limit? The households, impoverished by unemployment will continue to accumulate larger and larger money balances (in real terms)? Really? One would have thought that at some point the price level will have fallen far enough that the growing wealth embodied in money holdings will eventually induce additional spending. When it does, that can end Carlin and Soskice's deflationary spiral.⁴ (Since Carlin and Soskice do not have money in their model, this is invisible).

Similarly, I am sure cost-push inflation can occur in the sense that aggressive trade unions can sometimes extract large wage increases, and sometimes large numbers of unions can do it at the same time, and that this can lead to price increase. But if there is to be an *ongoing* inflation, we need a story about the other variables. Can it really be that V just increases to whatever level is necessary to validate a higher price level? Or must it be that the policymaker is creating, or anyway, permitting, an increase in M ?

And then there is QE. QE following the financial crisis was associated with changes in bank regulation which had the effect of causing banks to hold larger reserves, and that is one part of money demand. So an increase in money supply was partly matched by an increase in money demand. But QE following covid resulted in an increase in money holdings. For some reason, money demand increased and velocity therefore fell. Maybe there was unusually great concern about the future and particularly asset prices so that money became a more attractive asset, for example. But is there a sensible story that makes such a change permanent? If you think there is, relax (and tell me your idea). If you think the historical norms as to the demand for money will in the end reassert themselves, then you more or less have to think that pandemic QE is a cause of inflation. (On which, Congdon and Castañeda (2020)). So do you think inflation in 2022 and 2023 should be attributed to the war in Ukraine, or to the preceding increase in the quantity of money?

Later note: Now see Congdon (2024) for more on some of these themes.

⁴ This is the 'Pigou effect', or 'real balance effect', on which there is an IAQ.

References

- Carlin, W. and D. Soskice (2015). Macroeconomics: Institutions, instability, and the financial system. Oxford, OUP. BAL: 0765 e 177; SSL: HB172.5.CAR 2015; Union: 339 CAR
- Congdon, T. (2024). The quantity theory of money: a new restatement. London, Institute of Economic Affairs.
- Keynes, J. M. (1936). The General Theory of Employment, Interest and Money. London, Macmillan. Bal: 0775 e 001/01 and /02; Oxford Union: 330.156 KEY; SSL: HB171.KEY; History Faculty: M 060.6 Keyn;. <https://tinyurl.com/JMK36a>
- Radcliffe Committee (1959). Committee on the working of the monetary system (Radcliff Report) Cmnd 827. London, HMSO.
- Congdon, T. and J. E. Castañeda (2020). Inflation: The next threat? London, IEA. <https://iea.org.uk/publications/33536/>
- Friedman, M. (1956) The Quantity Theory of Money - A Restatement *in* Studies in the Quantity Theory of Money ed M. Friedman Chicago University of Chicago 3-21. Bal:0775 c 037; Soc:HG201.STU.
- Kaldor, N. (1970). "The new monetarism." Lloyds bank review (July): 1-17 Christopher Johnson (ed) 'Monetarism and the Keynesians', Lloyds Bank Annual Review, volume 4. Pinter, London. SSL: HG230.3.MON
- Keynes, J. M. (1936). The General Theory of Employment, Interest and Money. London, Macmillan. Bal: 0775 e 001/01 and /02; Oxford Union: 330.156 KEY; SSL: HB171.KEY; History Faculty: M 060.6 Keyn;. <https://tinyurl.com/JMK36a>
- Radcliffe Committee (1959). Committee on the working of the monetary system (Radcliffe Report) Cmnd 827. London, HMSO.