The Quantity Theory versus the Real Bills
Doctrine in Colonial America

Michael F. Sproul

Working Paper Number 775
Department of Economics
University of California, Los Angeles
Bunche 2263
Los Angeles, CA 90095-1477
January 16, 1998
msproul@csun.edu
Abstract

Colonial American currency has been cited as an example of the failure of the quantity theory. A case has been made in favor of the 'backing theory', which holds that money's value is determined by its backing, rather than its quantity. This paper explains that Colonial currencies were both backed and convertible, so that the quantity theory could not have been relevant to their value. The finding that currency values were best explained by their backing merely confirms the obvious proposition that backed money is worth its backing, but it neither refutes nor supports the quantity theory.
I. Introduction

A recent debate over the validity of the quantity theory concerns money and inflation in colonial America. In a series of articles, Bruce Smith (1985a, 1985b, 1988), found evidence from the colonial period in support of the theory that price levels are determined more by the backing of money than by its quantity.

When colonial currencies were carefully backed by future governmental surpluses, they held their value remarkably well. When such backing was not carefully provided, depreciation was the rule. The quantity of notes issued, on the other hand, bears little relation to currency values, or to colonial price levels. (Smith, 1985a, p. 156.)

Smith's conclusions were refuted by McCallum (1992). ¹

...the present discussion has featured nine episodes put forth by Smith as examples of the failure of classical monetary analysis. In each of these cases, large percentage increases in the stock of paper currency were followed by little or no response in price levels. Since the anticlassical position contends that specie supplies were minimal, it implies that large increases in real money balances occurred. The classical hypothesis, by contrast, is that outflows of specie (or commodity claims) occurred, with total real money balances remaining unchanged. The crucial implication is that, even at the episodes' peak years, colonies would be left with deflated levels of paper currency not significantly in excess of normal real money balances. (McCallum, 1992, p. 158.)

In what follows I hope to show that the evidence cited in this debate is not relevant to the theories in dispute. Smith's results do not refute the overall validity of the quantity

theory. They only indicate that it does not apply to the particular kind of money that he examined. McCallum's results can be interpreted in two ways. They could be taken as a defense of the quantity theory, but they are also consistent with Smith's hypothesis that the value of colonial American currencies was determined by backing.

II. Backed Money and the Real Bills Doctrine

A simple example will help to illustrate. Figure 1 represents a bank that has issued money ("Notes") in exchange for either gold or IOU's. (I have turned a conventional T-account on its side in order to emphasize that this bank's money is supported by its assets.) Assume initially that all notes are convertible on demand into one ounce of gold. If the bank begins by issuing 100 notes in exchange for 100 ounces of gold, then it is obvious that each note will be worth one ounce of gold. It is only slightly less obvious that the bank could then issue 100 more notes in exchange for IOU's worth 100 ounces of gold, and that this action would leave note values unchanged at one ounce each. The bank's IOU's, by assumption, could be sold for 100 ounces of gold, and in that event the bank would have 200 outstanding notes laying claim to 200 ounces of gold in its vaults.

\[
\begin{array}{c|c|c}
100 \text{ Notes} & +100 \text{ Notes} & \text{← Bank Liabilities} \\
100 \text{ oz. of gold} & +\text{IOU's worth} & \text{← Bank Assets} \\
& 100 \text{ oz. of gold} & \\
\end{array}
\]

Figure 1

The money I have just described is backed (by the gold and the IOU's) and it is convertible (into gold). The value of this money is not determined by the quantity theory.
of money, or by anything even resembling it. In point of fact, this money will be stable as long as the bank only issues new money in exchange for good security (i.e., for resources worth at least one ounce of gold). Economists will recognize this principle as the real bills doctrine, with two differences:

(1) Real bills adherents have historically maintained that “good security” means short-term commercial bills. In Figure 1, “good security” means anything worth 1 ounce of gold, so stocks, bonds, land, or lottery tickets would serve equally well.

(2) Some real bills adherents (e.g., Smith, 1776, p. 322) have held that the real bills rule would prevent inflation by assuring that the money supply moved in step with aggregate output. For the bank in Figure 1, the real bills rule prevents inflation by assuring that the money supply moves in step with its backing.

Now suppose we wish to use the money just described to test the validity of the quantity theory vs. the real bills doctrine. What would the test reveal? It would show that the value of this money depends completely upon its backing, and not at all upon its quantity. We would find, for example, that a 300% increase in the quantity of money (accompanied by a 300% increase in bank assets) would have no effect at all on the value of money. If the quantity of money were cut in half (while bank assets were also halved) then the value of money would still be unaffected. If the bank’s debtors were to default, so that, for example, the IOU’s previously worth 100 ounces of gold dropped in value to only 50 ounces, then even though the quantity of money had not changed, we would observe the notes falling below par, to $150/200=0.75$ ounces each. (The bank would either have to devalue its notes or suspend convertibility.) If the bank (or the government) printed 100 new notes and spent them, and the bank did not acquire additional backing, then the notes would fall to $100/200=0.50$ ounces each. In short, our test results would look just like Smith’s results for colonial American currencies.
What if the money issued by the bank in Figure 1 was not the only money in circulation, and that specie, foreign money, and commodity certificates were also commonly used? If we could form an estimate of the total circulating quantities of these various monies, we would probably find that when the bank in Figure 1 issued more notes, then comparable amounts of the other monies would be removed from circulation, leaving total money holdings unchanged. In short, we would find the same thing that McCallum found.

III. Backing, Convertibility, and Colonial Currencies

The finding that the quantity theory does not apply to backed money should not surprise anyone, but it does not mean that the quantity theory is invalid for fiat money. The real question is this: Was colonial American currency backed or was it fiat money? Both Smith and McCallum have neglected this point, and as a result have engaged in exactly the kind of fruitless debate that I have described above. Smith’s discussion of Maryland gives an obvious example.

The reason for focusing on Maryland derives from the nearly unique method adopted by that colony for backing its currency. Each of the colonies (at least ostensibly) backed its currency in some manner. Typically, currencies were backed either with future tax receipts or with mortgages (usually on land or metal plate). A time path for the value of this backing is generally impossible to obtain from existing data. However, Maryland backed the largest component of its note issues with the proceeds of a sinking fund invested in Bank of England stock. At preannounced dates (which were met in practice) some portion of the outstanding stock of these notes was to be converted into sterling (or, more precisely, sterling bills of exchange, described below) at a specified rate. (Italics added) Thus, a large component of Maryland paper money was a claim to future delivery of sterling. As the Sargent-Wallace view suggests that the value of money can be determined in essentially the same way as the value of privately issued claims, as Maryland notes were a claim against the sinking fund, and as there are fairly complete data on the market value of this sinking fund, a particularly appropriate setting is provided in
which to gather some empirical evidence on the Sargent-Wallace view. (Smith, 1985, pp. 1179-1180.)

Far from providing a "particularly appropriate setting...in which to gather some empirical evidence on the Sargent-Wallace view.", Maryland provides a setting in which the Sargent-Wallace view is tautological. The only thing we could possibly find by examining Maryland is that money which is backed and convertible is valued according to its backing and convertibility.

There are only two significant ways in which Maryland's money differs from the money described in Figure 1. First, Maryland's currency was not continuously convertible. Smith states (1985, p. 1199) that Maryland first introduced a paper currency in 1733, and that "One-third of the outstanding notes were to be redeemed in 1748 and the remaining two-thirds in 1764. These redemptions occurred as scheduled." Thus, Maryland's currency was inconvertible for 15 years at a time, after which it became briefly convertible. Of course, even the hypothetical bank in Figure 1 would not maintain truly continuous convertibility. Like all banks, it would close its doors at night and on weekends, thus making its notes temporarily inconvertible. Naturally, the notes would still be worth one ounce each over the weekend because of the expectation that convertibility would be restored on Monday. But if a bank can suspend convertibility for a weekend, it can suspend it for 15 years, and the only real difference between the two cases is that interest rates and uncertainty might play a more significant role when convertibility is suspended for a long period.

The second difference between Maryland and the bank in Figure 1 is that Maryland's notes were backed by a sinking fund which grew in size each year as earmarked taxes flowed in to the fund. Thus, while the bank in Figure 1 backs its notes with "IOU's worth 100 ounces", Maryland backed its notes with "Maryland's promise to levy future taxes sufficient to collect 100 ounces" (Figure 2).
Once this is understood, it is easy to see that there is no real difference between the two methods of backing illustrated in Figures 1 and 2. The market is capable of placing a numerical value on Maryland’s uncertain promises, just as it places a value on the IOU’s described in Figure 1. Since the difference between Figures 1 and 2 reduces to nothing more than different methods of collecting funds, we should not be surprised at Smith’s empirical results:

The coefficient on the money stock continues to be extremely small and highly insignificant. And finally, the coefficient on the (index of the) market value of the sinking fund is large, is significant at the 1 percent level, and has the theoretically predicted sign. In particular, increases in the market value of the backing for notes result in an appreciation in the value of Maryland currency. (Smith, 1985, p.1207.)

There is one additional difference between Maryland and the bank in Figure 1: Maryland’s sinking fund was denominated in British pounds, and it promised to convert into British pounds. Thus the assets of Maryland’s bank consisted entirely of IOU’s, rather than the mixture of gold and IOU’s described in Figure 1. It should be obvious, however, that the value of the bank’s money depends only upon the market value of the bank’s assets, and not upon whether those assets happen to be denominated in gold or in British pounds.
It remains to show that other colonial monies were also not fundamentally different from the money of Figure 1. Smith focuses on two types of colonial money: (1) bills of credit issued by loan offices, and (2) loan office notes issued by colonial land banks. The bills of credit were backed by future tax collections, just like the Maryland notes. Maryland funneled its taxes first into the sinking fund, which was then used to retire the notes, while other colonies used the taxes directly to retire their notes. Here again, the only difference between the backing of Maryland’s notes and that of other colonies was the number of hands through which the taxes passed on their way to retiring the notes. Thus, bills of credit issued by loan offices were backed (by future tax collections) and convertible (in the sense that a tax obligation of 1 British pound could be discharged with, for example, 1.5 Pennsylvania pounds).

Loan office notes issued by land banks could also be described by a diagram similar to Figure 1, the main difference being that the IOU’s would be backed by land that could be seized in case of default. Not surprisingly, Smith finds that when loan offices lent their notes prudently, for example, by lending no more than half the value of the property given in security, then note values were stable. On the other hand, if the property accepted as security were not adequate to repay the loan, or if legislatures extended undue time to delinquents, then a colony’s notes would fall in value.²

²Smith falls into an error on what constituted leniency on the part of the legislature:

Rhode Island was even more lenient regarding security for loans. Borrowers from Rhode Island often relent to others in Massachusetts, so that Rhode Island officials were obviously not sure what the ultimate backing of a loan was. (Smith, 1985, p. 550.)

If Rhode Island’s land bank lent 100 pounds to a Rhode Island resident who offered his land as security, and if he then relent the 100 pounds to someone in Massachusetts, the land bank’s claim to the original borrower’s land would in no way be compromised by the relending outside of Rhode Island.
A few peculiarities of land banks require an explanation. As Smith states:

...the notes issued by these land banks were *not* redeemable in commodities. They were typically given legal tender status, and colonial governments were obligated to accept them in payment of taxes. (Smith, 1985, p. 541.)

The fact that the notes were not convertible into *commodities* is misleading. If a colony lent 100 notes, which it declared to be legal tender at (say) one British pound per note, and if the land taken as collateral had a market value of (say) 200 British pounds, then colonists would be assured that each note could, in effect, be converted into something worth one British pound. When it was time to repay the loan that created the notes, the loan office would have a claim against borrowers equal to 100 British pounds plus interest. Thus, even though the notes were not continuously convertible, everyone knew that a time would come when each note could be used to pay off a debt of one British pound. ("...loan repayments would be made with notes accepted at par by the loan office." (Smith, 1985, p. 541.)) In effect, convertibility was delayed until the loan that created the notes came due. The delay in convertibility meant that uncertainty and interest rates would play a larger role in the notes’ value, but this is no different from the way the market values any other uncertain claim to future income.

To sum up: Smith focused on two types of money: (1) bills of credit, which were backed by future tax collections, and (2) loan office notes, which were backed by land. In spite of their superficial differences, both kinds of money were fundamentally the same as money that is backed and convertible. Thus, Smith’s finding that colonial currencies were valued according to their backing only confirms that backed money is worth its backing—a proposition that should not admit of dispute.
IV. McCallum’s Critique

In his critique of Smith, McCallum contends that specie flows were large enough to offset changes in the quantity of paper money. For example, when Smith found that New York tripled its paper money supply without experiencing inflation, McCallum replied that New York exported enough specie to leave its total money stock (paper plus specie) unchanged. By claiming that this evidence supported the quantity theory, McCallum only compounded Smith’s error. Smith’s mistake was in failing to point out that the money he examined could not possibly have been worth anything other than what its backing dictated. McCallum’s mistake was in trying to rescue the theory whose relevance should have been ruled out from the start. In fact, he strongly emphasizes that the quantity theory is not even the principle question in dispute.

A basic point is that the dispute between classical and anticlassical writers is not principally about the response of prices to money changes, but instead about the size of money stock changes associated with measures that pertain to paper currency alone. (McCallum, 1992, p. 157.)

McCallum’s statement is strikingly out of line with Smith’s view of the debate:

The primary result emerging from this examination is that all the colonies examined engineered extremely large (relative to typical government expenditures) note issues (reductions) that were not accompanied by inflation (deflation) or any depreciation (appreciation) of the notes issued against pounds sterling. (Smith, 1985, p. 533.)

While McCallum has a perfectly valid interest in the size of the money stock, we should not lose sight of the fact that the important question in this debate is indeed the response of prices to money stock changes. It is only because economists care about this issue that we are concerned about measurement of the money stock, currency flows, currency substitutes, etc.
V. Applications to Modern Currencies

The basic issue at stake is the validity of the quantity theory versus the real bills doctrine. Any debate should start by recognizing that when money is backed and convertible, then the real bills doctrine is correct and the quantity theory is inapplicable. When economists have attempted to implement the quantity theory, it has been in cases where governments have issued fiat money, which is, by definition, unbacked and inconvertible. If our ultimate goal is a theory that will correctly describe the value of money, then we must start by asking which type of money we are examining.

Suppose, for example, that we applied the methods of Smith and McCallum to the modern U.S. dollar. Our first observation would be that the dollar is inconvertible. We would then have to add that it has only been inconvertible to foreign governments since 1971, and to individuals since 1933. Since many governments have suspended convertibility only to resume it later, we must admit the possibility that the dollar is only temporarily inconvertible, and that it might not be a true fiat money. \(^3\)

Our second observation would take the form of a question: Is the dollar backed? Every modern textbook would answer that the dollar is not backed, because the dollar is not convertible into gold. But 'inconvertible' does not imply 'unbacked'. A bank can suspend convertibility for a weekend, but as long as it still holds collateral against its money over the weekend, the money is still backed. The Federal Reserve's own balance sheet currently identifies about $400 billion worth of gold and government bonds as "Collateral Held Against Federal Reserve Notes". Economists regard this as an accounting fiction because the Federal Reserve has suspended convertibility for many years. But as long as the

\(^3\)I have asserted in another paper (Sproul, 1997) that fiat money does not exist.
Federal Reserve still holds those assets, it could resume convertibility if it wanted to. And the fact that many central banks have restored convertibility after long suspensions makes this possibility more than idle speculation.

Now we have a question of real interest: Is the dollar a true fiat money (i.e., inconvertible and unbacked) or is it instead backed but temporarily inconvertible? This is the question to which the empirical methods of Smith and McCallum should be applied. If we found that the value of the dollar was best predicted by the dollar’s backing, then we would have evidence that the dollar really is backed, and that its value can best be predicted with the real bills doctrine. If, instead, we found that the quantity of money was the best predictor of the dollar’s value, then this would support the idea that the dollar is fiat money.

An empirical test of the real bills doctrine requires only that we estimate true market values of the Federal Reserve’s assets and liabilities. Note that we would not need an estimate of the amount of rival monies (specie, foreign money, etc.) or ‘derivative’ monies\(^4\) (checking accounts, credit cards, gift certificates, eurodollars, etc.) since none of these monies appear on the Federal Reserve’s balance sheet. We could also dispense with any estimates of money demand, velocity, real output, currency outside the country, etc., since backing and backing alone would be all that matters on real bills principles.

A proper comparison of the two theories would ask whether the value of the money issued by a specific entity was more accurately predicted by the assets and liabilities of that entity, or by some money demand function based on the quantity theory. In view of the

---

\(^4\)By ‘derivative money’, I mean money that is a claim to some other money, in the sense that a dollar in a checking account is a claim to one Federal Reserve note. By analogy, derivative shares of GM stock (options, warrants, hypothecated shares, etc.) are claims to a genuine share of GM stock. No economist would claim that an increase in the quantity of derivative shares of GM would affect the value of genuine shares of GM. Similarly, if the dollar is backed, we could not claim that an increase in the quantity of derivative monies would affect the value of the dollar.
limitations of colonial data, as well as McCallum's observation that new colonial money probably drove out comparable amounts of existing monies\textsuperscript{5}, we must conclude that the colonial period does not yield convincing evidence for one theory over the other.

Given the extensive data available on the assets and liabilities of modern central banks, as well as on various measures of the money supply, it is clear that a persuasive study would have to use modern data to discover whether the value of money is more accurately predicted by the quantity of money or by the backing held by the central bank. Furthermore, modern central banks back their money with bonds whose value can be precisely estimated. The advantage over Smith's use of "future government tax receipts in excess of expenditures" (1985, p. 533) as an estimate of backing is obvious.

Smith incorrectly claims that studies of modern central banks would be stymied by the abundance of fractional reserve intermediaries.

This similarity between theoretical specifications and colonial monetary arrangements reflects a simplicity of the colonial economy deriving from an absence of fractional reserve intermediaries. The absence of such intermediaries implies that it is unnecessary to decide whether private bank liabilities were money, and if so, to attempt to disentangle changes in this stock of high-powered money from changes in bank behaviour that might affect the 'money supply.' An attempt to conduct a study such as this one for any more recent period would certainly encounter problems of this sort. (Smith, 1985b, p. 535.)

\textsuperscript{5}McCallum's observation is open to question. He proposes that there were certain episodes when paper money was the only money in circulation:

...there are a few episodes where there is general agreement that little or no specie remained in circulation. Consequently, paper currency figures for these particular episodes can reasonably be taken as indicative of total money holdings. (McCallum, 1992, p. 149.)

McCallum's idea is that there were certain periods where Gresham's Law would have left only paper currency in circulation. However, Gresham's Law would operate only in a regime of fixed exchange rates, and exchange rates were not fixed, at least not officially. Since McCallum does not say exactly what he means by "general agreement", his claim is difficult to evaluate.
The trouble with this argument is that on real bills principles, the value of the dollar is determined only by the assets and liabilities of the Federal Reserve, and is unaffected by what I have called the 'derivative monies' issued in the private sector. Thus, a test of the real bills doctrine would not be hampered by fractional reserve intermediaries, since we need only look at the quantity of money issued by the Federal Reserve. A test of the quantity theory would have to examine derivative monies as well, and so Smith’s concern would apply with full force, but it is an open question whether the derivative monies of the modern period would present a greater complication than the various rival monies and derivative monies of the colonial period. In any case, quantity theorists can measure the money supply in any number of ways, and once this is done, the accuracy of the quantity theory can be directly compared to that of the real bills doctrine.

One additional point should be mentioned. Smith states that both the quantity theory and the Sargent-Wallace approach (which I equate with the real bills doctrine) fail to explain the experience of North Carolina.

Clearly then, the quantity theory cannot account for any of the North Carolina experience. How do we account for it according to the Sargent-Wallace approach? First, we should note that prior to 1748 there was no meaningful sense in which North Carolina backed its notes. The reduction in the money supply between 1715 and 1722 represents the only time prior to 1748 during which any notes were retired through taxation. Hence monetary expansions were not accompanied by increased future government revenue streams, and we should not be surprised by currency depreciation. Of course, since the quantity theory becomes a special case of the Sargent-Wallace view when money is unbacked, the failure of the quantity theory is also a failure of this viewpoint. Naturally, though, the Sargent-Wallace approach does no worse for this period than the quantity theory. (Smith, 1985, p. 1196.)

It is an overstatement to say that North Carolina’s notes were unbacked. A more accurate description is that they were backed by questionable assets (i.e., North Carolina’s
rather shaky promises to retire the notes) For example, each North Carolina pound may have been backed by North Carolina's promise to accept that pound in payment of future tax liabilities of one English pound. If that promise was likely to be broken, then people might have valued the North Carolina pound at only 0.1 English pounds. If the colony printed new notes (doubling their quantity, let us say) and spent them to cover revenue shortfalls, then the North Carolina pound would fall to 0.05 English pounds, assuming no change in the market value of the loan office's assets. The value of these notes would be entirely consistent with what their backing dictated, and therefore would not indicate a failure of the Sargent-Wallace approach (or of the quantity theory). The North Carolina case only shows that when the value of backing is unknown, then we can neither confirm nor refute the Sargent-Wallace approach. But we cannot call that approach a failure when we have no evidence one way or the other.

VI. CONCLUSION

Smith has found that the value of colonial currencies was more accurately predicted by their backing than by their quantity. He views this as empirical confirmation of the backing theory of Sargent and Wallace, and as a refutation of the quantity theory. However, if colonial currencies were backed and convertible, then Smith's findings only demonstrate the obvious fact that backed money is worth its backing. Smith's findings also cannot be viewed as a refutation of the quantity theory, since the quantity theory is not applicable to backed money.

In his critique of Smith, McCallum claims that Smith erred in his measurement of the money supply, and that currency values were best predicted by the quantity theory. McCallum's measurement techniques are subject to some reservations, but the crucial
point is that if colonial currencies were in fact backed, then the quantity theory would not be relevant.

All colonial currencies were, in fact, backed and convertible. Convertibility was often delayed or uncertain, and backing was often questionable, but this is true of all securities. The proposition that colonial currencies were backed implies that their value would be correctly predicted by a theory that closely resembles the real bills doctrine.

If we find that the value of a certain kind of money is best predicted by the real bills doctrine, then that is evidence that the money in question is backed. If we find that the money's value is best predicted by the quantity theory, then that is evidence that the money in question is fiat money. In neither case would we have evidence for the overall validity of one theory or the other, since each theory only applies to a particular kind of money. The colonial period, with its sparse data on backing and on various measures of the money supply, does not yield persuasive evidence for either theory. Modern data is much better suited to the task, but economists should begin their studies by questioning the existence of fiat money, rather than using it as a starting point.

REFERENCES


Smith, Bruce D. "American Colonial Monetary Regimes: The Failure of the Quantity Theory and Some Evidence in Favour of an Alternative View." Canadian Journal of Economics volume 18, August 1985, pp. 531-65. (a)

Smith, Bruce D. "Some Colonial Evidence on Two Theories of Money: Maryland and the Carolinas", Journal of Political Economy, volume 93, number 6, December 1985, p. 1178. (b)


