Economies of Scale in Cash Balances Reconsidered

Karl Brunner
University of California - Los Angeles

Allan H. Meltzer
Carnegie Mellon University, am05@andrew.cmu.edu

Follow this and additional works at: http://repository.cmu.edu/tepper
Part of the Economic Policy Commons, and the Industrial Organization Commons

Published In
The Quarterly Journal of Economics, 81, 3, 422-436.

This Article is brought to you for free and open access by Research Showcase @ CMU. It has been accepted for inclusion in Tepper School of Business by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.
Economies of Scale in Cash Balances Reconsidered

by

Karl Brunner and Allan H. Meltzer

1967

Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

(Continued on inside back cover)
ECONOMIES OF SCALE IN
CASH BALANCES RECONSIDERED *

KARL BRUNNEE AND ALLAN H. MEITZER

The Baumol model, 423.—The Tobin model, 427.—Conclusion, 435.

Renewed interest in the demand for money has resulted in the clarification of a number of issues and has focused attention more sharply on some others. One subject of dispute has been the existence of economies of scale in the demand function for money by business firms. A familiar version of the quantity theory of money, \( M = kY \), suggests the absence of economies of scale. Our own approach, the wealth adjustment model, reaches a very similar conclusion. It implies that — to a first approximation — the demand for money estimated from cross-sections of business firms is linear in the logarithms and unit elastic with respect to sales or transactions. This proposition is supported by a substantial body of evidence from cross-section and time series regressions.¹

The wealth adjustment model has not yet been derived from a more general theory of wealth or utility maximization or from a production function. This is of particular importance since Baumol has claimed — contrary to the quantity theory — that rational behavior implies that the demand for money by business firms will increase less than proportionally to the increase in the volume of transactions.² The asserted conflict between the Baumol analysis and the quantity theory or the wealth adjustment model raises a number of questions: Are business firms irrational in the management of their cash balances? Do the quantity theory of money and the wealth adjustment model require irrational behavior? Or, has Baumol drawn an incorrect or overstated inference from his theory?

A number of economists have attempted to find evidence for

* We gratefully acknowledge research support from the National Science Foundation, the University of California, Los Angeles, Carnegie Institute of Technology, and the University of Chicago.


2. W. J. Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," this Journal, LXVI (Nov. 1953). Note especially p. 530 where he argues that either the quantity theory is not based on rational behavior or it "cannot have general validity."
Baumol's conclusion or for what is said to be a similar conclusion by Tobin or have disputed the evidence from the cross-section studies. While we find none of these studies sufficiently convincing to cast serious doubt on the empirical results we have obtained from tests of the wealth adjustment model, it is useful to show that those who seek to support Baumol's conclusion — that cash balances are subject to sizable economies of scale — should not expect to do so on the basis of his theory. Indeed, we will reach the conclusion that Baumol's model does not imply important economies of scale in the demand for cash balances when it is completed along the lines that he sketches. A somewhat similar conclusion applies to the Tobin hypothesis. Neither the Baumol nor the Tobin formulations are inconsistent with the use of the quantity theory of money (or the wealth adjustment model) as a first approximation to the cross-section demand for money by business firms.

THE BAUMOL MODEL

Baumol derives the transaction demand for money equation by minimizing the cost of holding cash balances at given interest rates and transaction costs. It is clear from his presentation, though not always recognized, that three separate cases are considered. In the first two, firms cannot receive cash by selling output. Cash is obtained only by borrowing or by selling financial assets. In these two restricted cases, the familiar square root formula for optimal cash balances is obtained. But it should be noted that difficulties

3. James Tobin, "The Interest Elasticity of the Transactions Demand for Money," Review of Economics and Statistics, XXXVIII (Aug. 1956). Tobin is concerned primarily with the interest elasticity of transaction balances and not with the question of economies of scale. However, Tobin states (p. 247) that his equation will produce "essentially the same results as Baumol's" square root formula or the more complete Baumol hypothesis considered here. This statement misled one author of this paper (Meltzer, "A Cross-Section Study," op. cit.) and has apparently misled others into the belief that the Tobin equation suggested the importance of scale economies in cash balances.


5. For a discussion of some of the problems of applying the lot-size
occur when an attempt is made to specify the time period to which the analysis pertains or to aggregate over several firms. The simple model is an incomplete statement of individual firm behavior and is unsuitable for aggregation. The problem is that firms are not permitted to receive money from sales or payments from accounts receivable, a statement that is unlikely to be true if the analysis is applied to more than a short time span or to more than a single firm. To put the matter succinctly, it is most likely that the larger the number of firms over which one aggregates, the smaller the time periods to which the particular conclusion of Baumol's analysis pertains. Even for a single firm, the time period during which there are no receipts from sales or receivables is likely to be no more than a few days and certainly less than the monthly or quarterly dates at which we observe cash balances. To avoid the problem, some statement must be made about the period to which the simple model applies or about the money holding behavior of firms during periods in which there are cash receipts from sales.

Baumol does not overlook the latter case, although many others (including the second-named author of this paper) have done so when discussing his conclusion. However, he concludes that a similar transaction demand function is obtained if firms are permitted to receive cash at the beginning of a transaction period. Specifically, he obtains for this case

\[
R = C + T(k_w + k_d)/i
\]

where

- \(R\) = the number of dollars of receipts from sales that are not invested in securities
- \(C\) = the amount withdrawn from investment balances (dollars)
- \(T\) = the number of dollars (actual or expected) used to make payments during the period between receipts
- \(i\) = the interest rate on a loan or the opportunity cost of holding money rather than financial assets
- \(k_w\) = the marginal cost of withdrawing cash
- \(k_d\) = the marginal cost of depositing cash.

Earlier, he had obtained the familiar result that

\[
C = \frac{\sqrt{2b_wT}}{i} \quad (b_w = \text{the fixed cost of withdrawing cash})
\]

formula to cash balances see W. Beranek, "Some Implied Assumptions of Standard Inventory Models," forthcoming.

6. Baumol, op. cit., p. 549. Baumol obtains this equation by minimizing a cost function with respect to \(I\), the number of dollars invested from the proceeds of sales, and with respect to \(C\), the number withdrawn (or withheld) from investment balances.
applicable to the period in which there are no receipts from sales. Since part of the money received is held in cash and since he assumes no additional receipts, equation (2) may be substituted into the formula for $R$ to obtain, as Baumol suggests:

$$R = \sqrt{2b_T T/i} + T(k_r + k_d)/i.$$  

This procedure raises two questions of interest. First, why do firms ever hold “speculative” balances? Second, what is the amount of the cash balance that we should expect to observe if we apply the Baumol analysis to a sample of firms? We consider each of these questions in turn.

Baumol notes that “any receipts exceeding anticipated disbursements will be invested, since, eventually, interest earnings must exceed (‘brokerage’) costs of investment.” This statement is difficult to reconcile with his later observation that suggests the existence of speculative balances. Either money balances are identical with transaction balances or Baumol’s model is incomplete and requires supplementary statements to describe the division of money balances between speculative and transaction balances. Moreover, his discussion is in terms of known receipts and expected disbursements, a probabilistic statement that seems to include the precautionary motive as well. Despite Baumol’s statements to the contrary, the analysis appears to merge a firm’s “transaction balances” with total money balances. Our evidence suggests that the merger is appropriate, that better results are obtained if money balances are not separated according to “motive.”

Turning to the amount of cash that we would expect to observe, on Baumol’s assumptions, we must distinguish two separate components of the average cash balance. One is the amount that optimizing firms will withhold from investment. During a part of the transaction period, with length $R/T = (T - I)/T$, optimizing firms will hold an average amount $R/2$ in cash. From equation (1), above, it is clear that this amount includes a component $C$ that is described as the amount initially withheld from investment. The second component is the average amount, $C/2$, held during the period $I/T$ in which there are no receipts from sales. The average money balance for a firm must then be the weighted average of $C/2$ and $R/2$ with weights given by the proportion of the period

7. Ibid., p. 547.
8. See also the discussion in Baumol’s fn. 2, p. 546, where various risks are incorporated into $b$ and $i$.
during which we should expect to observe these balances if the Baumol analysis is used to determine a firm's cash balance. We denote the weighted average balance by $M$:

$$M = (R/2) [(T - I)/T] + (C/2) (I/T).$$

Substituting for $R$ from equation (1), and recalling that $R = T - I$, we obtain

$$M = C/2 + (R/2i) (k_w + k_d).$$

Again substituting for $R$ from equation (1) and rearranging terms, this becomes

$$M = \sqrt{b_wT/2i} \left[ 1 + (k_w + k_d)/i \right] + T/2 \left[ (k_w + k_d)/i \right]^2. \quad (4)$$

It should be clear that the often used square root formula is a particular case of equation (4), the case in which $k_w$ and $k_d$ are zero. As Baumol notes, this assumption becomes troublesome once firms are permitted to receive cash from sales or repayment of accounts receivables. If the marginal cost of investing and withdrawing cash is zero, it pays to invest in non-cash assets at any positive interest rate and for any arbitrarily short period of time. Thus the familiar square root formula for optimal money balances cannot be obtained under these conditions since we cannot assume $k_d$ and $k_w = 0$. Instead, the Baumol assumptions imply the average (optimal) balance is a quadratic function of the volume of transactions.

However, we should note that equation (4) is not necessarily inconsistent with the quantity theory of money as an explanation of a firm's demand for money. Two cases may be distinguished. First, if the fixed cost of withdrawing cash, $b_w$, is zero, the square root term vanishes, and the desired balance becomes a linear function of the volume of transactions with a coefficient depending on the marginal cost of depositing and withdrawing cash and the interest return from holding financial assets. The firm's cash balance increases linearly with the volume of transactions, interest rates and marginal costs remaining unchanged. Moreover, for small values of $b_w$, the quantity theory remains a reasonable first approximation as the evidence to which we have referred seems to suggest.

Second, for a given value of $b_w$, the existence of economies of scale is unlikely to be detected in the cash balances of large firms behaving according to the Baumol model of rational behavior. The elasticity of $M$ with respect to $T$, denoted $\varepsilon(M, T)$, is

$$\varepsilon(M, T) = \frac{[1 + (k_w + k_d)/i] \sqrt{b_w T/2i} + [(k_w + k_d)/i]^2 T/2}{2 [1 + (k_w + k_d)/i] \sqrt{b_w T/2i} + [(k_w + k_d)/i]^2 T}. \quad (5)$$

2. Ibid., fn. 6, p. 549.
As $T$ approaches zero, the elasticity approaches one-half, the result obtained from the simpler Baumol model (equation 2) for all values of $T$. But for large $T$, the elasticity in equation (5) approaches one, the value expected from the quantity theory. Once receipts of cash from sales or accounts receivables are incorporated in the model, the value of the elasticity increases toward unity with increases in the volume of transactions. Unless $b_\omega$ is extremely large, the elasticity approaches unity quite rapidly as $T$ rises, since $(k_\omega + k_d)/i$ is likely to be a small fraction.

However, speculation about the values of the parameters is unnecessary. Parameter estimates have been reported for regressions based on industry subaggregate and individual firm data. These data suggest that economies of scale are quite minor and hard to detect for firms with transactions or sales greater than $\$25$ or $\$50 million per year, even if they occur up to that point. Moreover, the data reject the often repeated implication of the simpler Baumol hypothesis since they suggest that, where there are economies of scale in holding cash balances, the economies are exhausted at relatively low values of transactions volume.

Thus the Baumol model provides little reason for abandoning the earlier conclusion that to a first approximation the quantity theory explains the observed cash holdings of business firms when the observations are from cross sections at a point of time. It is only when we are concerned with the behavior of relatively small firms or the distribution of cash balances within industries that the economies or diseconomies of scale implied by the Baumol model may become apparent. The difference between the cash balances expected under the Baumol model and the quantity theory is sufficiently small that, even if the Baumol model is correct, it provides little basis for rejecting the quantity theory or concluding that sizable economies of scale are to be expected.

**THE TOBIN MODEL**

In Tobin's approach the firm maximizes the net revenue from total transaction balances assumed to consist of bonds and money.

3. The elasticity of cash with respect to sales was computed for fourteen industry groups in each of nine years. The cross sections are based principally on data from *Statistics of Income*, U.S., Internal Revenue Service. The mean value of the elasticity for each ranges from 1.01 to 1.16. The mean of the 126 regressions is 1.04. Eighty per cent of the computed elasticities are above unity. For further details and tests of alternative functions, including the quadratic, see Meltzer, "A Cross-Section Study," *op. cit.*

Transaction balances are said to be held as a means of bridging the gap between receipts that occur at discrete intervals and continuous expenditures. Tobin does not specify the determinants of total money holdings or of the distribution of money balances between the transaction and nontransaction components. His principal concern is to show that interest rates will affect the amount of transaction balances held in the form of money.

Nevertheless, Tobin's analysis has been used to suggest that there may be substantial economies of scale in transaction balances of aggregate or individual firms. In this section, we show that to a first approximation each firm, operating according to the Tobin model, should regard its demand function for (transactions) money as unit elastic in the volume of transactions. We then discuss some problems that arise if the Tobin model is applied to firms in the aggregate. Only those portions of Tobin's analysis that are required by the present discussion will be reproduced.

Tobin solves for the average amount of transactions balances that is held in bonds rather than money. This amount is

$$\bar{B} = \frac{(n - 1)}{2n} \left[ Y \left( 1 - 4b^2/r^2 \right) \right] n \geq 2, r \geq 2b$$

where

- $\bar{B}$ = the average amount of bonds held
- $n$ = the number of transactions made to exchange money for bonds (restricted to integer values)
- $Y$ = the value of receipts at the beginning of the period (equal to the amount disbursed during the period)
- $b$ = the variable cost of exchanging money for bonds or bonds for money
- $r$ = the interest rate paid on bonds.

He had previously defined

$$\bar{C} = Y/2 - \bar{B} \quad \text{($\bar{C}$ = the average cash balance)}$$

so that

$$\bar{C} = \frac{Y}{2} \left[ 1 - \frac{(n - 1)}{n} \left( 1 - 4b^2/r^2 \right) \right] n \geq 2, r \geq 2b.$$  

We note, first, that up to the point at which it pays to make two transactions, all receipts ($Y$) are held in the form of cash; cash balances increase in direct proportion to the volume of transactions. $\bar{C} = (1/2) Y$. Second, as $n$ becomes relatively large, cash balances increase in direct proportion to the volume of transactions (or receipts) given $b$ and $r$. Third, for constant values of $n$, the same conclusion is reached. But it is an essential part of the Tobin

5. Ibid., p. 245.
analysis that \( n \) is variable. The existence of economies of scale in the holding of cash thus appears to result solely from the effect of variations in \( n \). This effect will be examined presently.

Before doing so, it is useful to investigate the expected value of the transactions elasticity of cash balances, denoted \( \varepsilon(C, Y) \), ignoring the effect of \( Y \) on \( n \). Differentiating equation (6) shows that

\[
\frac{\partial C}{\partial Y} = \frac{1}{2} \left[ \frac{1}{n} - \frac{n - 1}{n} \left( 1 - \frac{4b^2}{r^2} \right) \right]
\]

and that \( C/Y = \partial C/\partial Y \) so that \( \varepsilon(C, Y) \) equals unity under the Tobin analysis, just as it would under the quantity theory. Note that \( \varepsilon(C, Y) \) is independent of \( n \), the number of exchanges of bonds for money. Thus, we are led by the Tobin analysis, to expect neither economies nor diseconomies of scale in the holding of money balances, if changes in \( n \) are ignored.

There remains the problem of investigating the importance of scale economies when \( n \) is permitted to vary. Tobin makes the optimal number of transactions, \( n^* \), dependent on \( Y, r, b, \) and \( a \) (\( a \) = the constant fixed cost component of transactions costs); \( n^* \) increases directly with \( Y \) and \( r \) and inversely with \( a \) and \( b \). A solution for \( n^* \) can be obtained from Tobin's analysis. Optimal \( n (= n^*) \) is found when the marginal net revenue is not less than the (constant) marginal cost of increasing the number of asset exchanges, i.e., when

\[
(7) \quad n^*(n^* + 1) \geq Yr\left(1 - \frac{2b}{r}\right)/2a.
\]

However, Tobin restricts \( n \) and \( n^* \) to integer values to maintain the equality of cash receipts and expenditures for the transaction period. It is clear, therefore, that for any given values of \( a, b, \) and \( r \) there will be no change in \( n^* \) unless \( Y \) changes by an amount sufficient to raise the optimal number of transactions above the preexisting \( n^* \). Between the points at which these discrete changes take place, cash balances are unit elastic with respect to the volume of transactions as our previous analysis implies. Each time the optimal number of transactions increases by one unit, the optimal cash balance shifts down by a finite amount dependent on \( r \) and \( b \). The relation is again unit elastic until the next point at which the optimal number of transactions increases. This relation is represented in Figure I.

The extent to which an individual firm's demand for money will depart from unit elasticity depends (in the Tobin analysis) on the frequency with which a firm observes changes in receipts or expenditures that are sufficiently large to cause the firm to cross one of the

"jump" points at which \( n^* \) increases. Tobin is not directly concerned with the problem of economies of scale, does not solve for the cash balance, and does not mention the jump points in optimal cash balances that arise from the restriction on the optimal number of transactions to an integer value per transaction period. But his analysis can be used to compute the amount by which receipts must change before there is a change in the optimal number of transactions between money and bonds. He shows that the number of asset exchanges does not exceed two per period unless the requirements for his Case IV are met, i.e., until

\[
(1/12) Y r (1 - 2b/r)^2 \geq a. 
\]

From equation (7) we obtain the amount by which receipts must change \((\Delta Y)\) to raise or lower the optimal number of transactions. For given \( a, b \) and \( r \), the condition is

\[
\Delta n^* (2n^* + \Delta n^* + 1) \leq \Delta Y (r/2a) (1 - 2b/r)^2. 
\]

A jump point is crossed if \( \Delta n^* = 1 \), i.e., if

\[
\Delta Y \geq [4a(n^* + 1)]/r(1 - 2b/r)^2. 
\]

Assume that a firm has reached the minimum level of receipts that satisfies equation (8). For \( n^* \) to increase by one unit, the firm must experience a change in receipts

7. Ibid., p. 245.
ECONOMIES OF SCALE IN CASH BALANCES

\[ \triangle Y = \left(\frac{1}{3}\right) (n^* + 1) Y_{\text{min}} \]

where \( Y_{\text{min}} \) is the minimum value of \( Y \) at which \( n^* = 2 \).

When \( n^* = 2 \), \( \triangle Y = Y_{\text{min}} \). Unless receipts double from one transaction period to the next, a firm that is just above the threshold level of sales at which \( n^* \) becomes 2 need not be concerned with the possibility that \( n^* \) may become 3. And a firm that is midway between \( n^* = 2 \) and \( n^* = 3 \) must experience a 50 per cent increase in sales before it crosses another jump point and experiences additional economies of scale.

Table I shows the variation in receipts per transaction period that is consistent with an unchanged value of \( n^* \). The values in the table are obtained from equation (9) and are used to mark the “jump points” in Figure I. Each time \( n^* \) increases by one unit, the distance between jump points increases by \( 1/3Y_{\text{min}} \). The minimum value of \( Y \) that is the lower bound for each successive value of \( n^* \) increases also, so that the change in receipts required to move a firm from one optimal \( n \) to the next becomes a smaller proportion of the level of receipts that mark the entrance to a particular range. Nevertheless, column (3) of the table shows that the percentage change in receipts per period consistent with retaining a constant value of \( n^* \) remains relatively large as \( n^* \) increases.

A continuation of the table would more clearly reveal the pattern, viz., as \( n^* \) doubles, the values in column (3) are reduced by 50 per cent. But this information tells us only that each firm that is not close to a jump point may treat \( n^* \) as a constant up to relatively large values of \( n^* \) and proceed on the assumption that cash balances are unit elastic in the volume of transactions. The table

<table>
<thead>
<tr>
<th>Value of ( n^* ) (1)</th>
<th>Minimum value of ( Y ) at lower bound (2)</th>
<th>Percentage of lower bound by which receipts may change without increasing ( n^* ) (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 ( Y_{\text{min}} )</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>3 ( 2Y_{\text{min}} )</td>
<td></td>
<td>( 2/3 = 66.7% )</td>
</tr>
<tr>
<td>4 ( 10/3Y_{\text{min}} )</td>
<td></td>
<td>( 1/2 = 50% )</td>
</tr>
<tr>
<td>5 ( 5Y_{\text{min}} )</td>
<td></td>
<td>( 6/15 = 40% )</td>
</tr>
<tr>
<td>6 ( 7Y_{\text{min}} )</td>
<td></td>
<td>( 1/3 = 33.3% )</td>
</tr>
</tbody>
</table>

\( Y_{\text{min}} \) is given by equation (8) of the text; \( Y \) is computed from equation (9); transactions costs and interest rates \( (a, b, r) \) are held constant.
provides no basis for concluding that economies of scale are exhausted at low values of \( n^* \).

The problem is that the Tobin model gives no information from which the length of the transaction period (or the value of \( Y_{\text{min}} \)) can be obtained. But the length of the transaction period plays a critical role in the interpretation of the analysis and particularly so for the question of economies of scale. Without information about the length of the transaction period, nothing can be said about the size of scale economies. For example, if the transaction period is a week, a firm that makes five asset exchanges probably exhausts most of the economies of scale that may exist. Such a firm engages in securities transactions every day. Even if receipts vary by more than the 40 per cent per week required to move the firm from the lower bound at which \( n^* = 5 \) to \( n^* = 6 \), few additional economies of scale can be expected at the increased \( n^* \). This conclusion does not follow if \( n^* \) remains at five and the transaction period is a month. Moreover, the conclusion does not follow if the firm makes semiweekly transactions (\( n^* = 26 \)) and the transaction period is a calendar quarter. Additional economies of scale would be expected at values of \( n^* \) larger than 26 if the transaction period is three months. And smaller percentage changes in sales would be required to move the firm’s receipts into the range at which the additional economies could be realized.

However, as \( n^* \) increases additions to \( n^* \) produce smaller downward shifts in the curve at each of the successive jump points. This conclusion is independent of any assumption about the length of the transaction period. A similar conclusion holds for the average cash balance. But the absolute size of the reduction in the average cash balance cannot be obtained from the analysis Tobin presents. Assumptions must be introduced about the values of \( b \) and \( r \). However, Figure I suggests the relative size of the reductions in log \( C \) as \( n^* \) changes.

Table II shows the computed values of the elasticity and the average cash balance for successive ranges of the curve and for one set of assumed values: \( b = .01, r = .04 \). The elasticities shown in the table are minimum values for \( n^* = 2 \ldots 7 \) and arbitrarily chosen values of \( b \) and \( r \). The computations are based on the assumption that a firm moves along a curve connecting the cash balance held when the receipts from sales are the minimum amount at which \( n^* = 2 \) to the cash balance that is held when receipts are just sufficient to make \( n^* = 3 \), etc. The Tobin analysis denies that this is the path that the firm follows and implies instead that a firm
moves between jump points along a curve that is unit elastic. If the
Tobin hypothesis is correct, values of the elasticity computed by
the method used to construct the series in the table will be under-
estimates of the elasticities obtained from econometric studies for
all values of \( 0 < b < 2r \). The size of the underestimate depends on
the values of \( b \) and \( r \). But whatever the values chosen, the minimum
values of the elasticities approach unity as \( n^* \) increases. It is likely,
therefore, that even if the Tobin hypothesis is true, it may be
difficult to distinguish between this model and the quantity theory
in cross-section regressions. The data in the table suggest that the
elasticity computed by regression will be close to unity if the true
values of \( b \) and \( r \) are in the neighborhood of those we have chosen.

### TABLE II

VALUES OF THE AVERAGE CASH BALANCE AND OF THE MINIMUM
ELASTICITY OF CASH BALANCES BETWEEN VALUES OF \( n^* \)

\[
\begin{array}{cccccc}
\hline
n^* & Y_{min} & \frac{5}{10}Y_{min} & \frac{5}{10}Y_{min} & 5/10Y_{min} & 5/10Y_{min} \\
\hline
2 & Y_{min} & \frac{5}{10}Y_{min} & \frac{5}{10}Y_{min} & 5/10Y_{min} & 5/10Y_{min} \\
3 & 2Y_{min} & \frac{3}{1}Y_{min} & \frac{3}{1}Y_{min} & \frac{3}{1}Y_{min} & \frac{3}{1}Y_{min} \\
4 & 10/3Y_{min} & 5/15Y_{min} & 5/15Y_{min} & 5/15Y_{min} & 5/15Y_{min} \\
5 & 5Y_{min} & Y_{min} & Y_{min} & Y_{min} & Y_{min} \\
6 & 7Y_{min} & \frac{63}{48}Y_{min} & \frac{63}{48}Y_{min} & \frac{63}{48}Y_{min} & \frac{63}{48}Y_{min} \\
7 & 28/3Y_{min} & \frac{5}{3}Y_{min} & \frac{5}{3}Y_{min} & \frac{5}{3}Y_{min} & \frac{5}{3}Y_{min} \\
\hline
\end{array}
\]

* Computed from \((\Delta C/C)Y/(\Delta Y)\).

We may summarize our analysis up to this point by noting that
the Tobin model suggests that each firm that is not close to a jump
point may assume that its cash balance is unit elastic with respect
to the volume of transactions. We noted earlier that the same con-
clusion holds when the volume of transactions is relatively large or
when it is so small that \( n^* \) is less than 2. To a first approximation
every firm may consider its demand function to be unit elastic. The
error in the approximation will be quite small in general and will
decline with the number of exchanges between bonds and money.
In short, there is little conflict between the implications of the
Tobin analysis and those of the quantity theory when both are
applied to individual firms.

While the conclusion just reached is valid for each firm indi-

g individually, we have found that it need not hold for the aggregate of
firms or households under the Tobin analysis. Tobin provides no
cues that would help to determine the expected size of the reduct-
in average cash balances if large and small firms are compared.
Of particular importance is the absence of a statement about the length of the transaction period, the time between cash inflows. Without this information, we do not know whether the transaction period is assumed to be the same for firms of different size, so we cannot compare observed average cash balances. Moreover, we do not know the value of \( b \), so we cannot compute the magnitude of the economies of scale between size groups from equation (6).

However, the data from Statistics of Income can be used in one of two ways to check the implication of the Tobin analysis and to bring out the problem in interpretation. Equation (6) implies that the ratio of cash to the volume of transactions per transaction period should be 1/2 for firms with the smallest volume of transactions, since such firms may be assumed to make no asset exchanges. (1) Assume that the length of a transaction period is the same for all firms. If there are economies of scale, the ratio of cash to sales should fall with size of firm. During the years 1955–61, the ratio of cash to sales was lower for (total) manufacturing firms of smallest size (less than $25,000 in assets) than for any size class up to the largest (assets in excess of $250 million). Firms in the top asset size class have lower cash/sales ratios than those in the bottom size class in only two of the seven years. (2) Alternatively, assume that the length of the transaction period differs among firms in different size classes. The data for 1955–61 suggest that the average ratio of cash to annual sales for firms of smallest size is .038. For firms of largest size, this ratio is .042. The data can be used with equation (6) to illustrate the effect of economies of scale on the length of the transaction period. If there are no exchanges of bonds for money or of money for bonds, the ratio of cash to receipts (or sales) per transaction period is .50. Since firms of smallest size may be assumed to make no asset exchanges, a transaction period of approximately 4 weeks (.50/.038) is required to make the data consistent with the Tobin hypothesis. If firms of largest size make no asset exchanges per transaction period, the ratio of cash to receipts per transaction period is .50 for them also, and their transaction period is 1 month (.50/.042), slightly longer than the transaction period for firms of smallest size. Inspection of equation (6) shows that any increase in the number of asset ex-

---

changes reduces the ratio of cash to receipts per transaction period and thus increases the length of the transaction period required to make the Tobin hypothesis consistent with the annual data.

If we are unwilling to accept the finding that the largest firms receive cash less frequently than the smallest firms, one or two conclusions must be rejected: (1) the Tobin model implies economies of scale; or (2) Tobin's model is applicable to the cross-section observations. The data suggest that it is the latter conclusion that must be rejected. The ratio of cash to receipts per transaction period (.50) may be assumed constant \( n^* = 0 \) for firms with total assets of less than $25,000. For the data to be consistent with the Tobin model, the transaction period for small firms must be variable since the ratio of cash to annual sales is variable. It is not unlikely that the transaction period for firms in other size classes varies also. Variations in payments schedules in turn suggest that firms search for an optimum payments schedule and that the determination of optimum money balances is obtained as part of a larger optimizing problem.

The problem may be only that we have pushed the analysis too far. Tobin was interested primarily in showing that an individual or firm's demand for transaction balances is (beyond a minimum value of transactions) dependent on interest rates. He stressed neither the aggregate relationship nor the importance of economies of scale. In any case, his analysis does not seem to be consistent with the cross-section observations for industries unless some additional statements are introduced to explain differences in the length of transaction periods for firms that differ in size. And his analysis does not suggest the importance of economies of scale for individual firms with the perhaps minor exception of those that are close to a "jump" point.

**Conclusion**

There is probably little need to point out that the existence of economies of scale in the demand for money is an empirical question. We have not been primarily concerned with that question here since the evidence suggesting that scale economies are minor or nonobservable has been presented elsewhere. The issue has been whether or not "rational behavior" implies that there should be economies of scale in the cross-section demand for money by firms or individuals or in the demand for money by individual firms over time.
It has now been established that rational behavior of the Baumol or Tobin type does not deny that the quantity theory of money explains the cross-section demand for money by firms (at least the transaction component) to a first approximation. This, of course, does not mean that the quantity theory provides a good explanation; both models may be false. But now the circle is complete, since we started with the observation that, to a first approximation, the quantity theory explained the cross-section observations quite well.

Thus it appears that for an individual firm, the Baumol and Tobin models are not alternatives to the familiar quantity theory. They both imply it and as such perhaps furnish a firmer foundation for it. The same conclusion holds under the Baumol model for firms of different size at a particular point in time, while the Tobin model is not sufficiently specified to permit a conclusion to be drawn for the aggregate.

Of course, nothing that we said here should suggest that economies of scale are inconsistent with some other model based on rational behavior that might be constructed. But analysis and evidence seem to indicate that an extended search for substantial economies or the construction of theories that imply them will not prove fruitful, if the assumption of fixed payments schedules is retained.

Ohio State University
Carnegie Institute of Technology


