A Monetary Perspective on Underground Economic Activity in the United States

This article was prepared by Richard D. Porter and Amanda S. Bayer of the Board's Division of Research and Statistics. Footnotes appear at the end of the article.

A growing underground economy in the United States and in other countries has been widely reported in recent years. The underground economy is thought to reflect efforts to evade taxes and government regulation. Although no single definition of such activity has been universally accepted, the term generally refers to activity—whether legal or illegal—generating income that is either underreported or not reported at all. Some investigators narrow the definition to cover only income produced in legal activity that is not reported in the national income statistics.

Discussion of underground economic activity intensified in the late 1970s with the publication of two estimates, derived from aggregate monetary statistics, of the size of the underground economy in the United States, one by Peter Gutmann and the other by Edgar Feige.[1] Since then, numerous estimates have been made of the scope of this sector in the United States and in other countries. The magnitude of some of these estimates has occasioned congressional hearings and various government studies. In 1979, the Internal Revenue Service (IRS) estimated that for 1976, individuals failed to report between $75 billion and $100 billion in income from legal sources and another $25 billion to $35 billion from three types of illegal activity—drugs, gambling, and prostitution.[2] In 1983, the estimates of unreported income from legal sources for 1976 were raised to $131.5 billion while the estimates of income from illegal sources dropped to $13.4 billion.[3]

In this more recent study, the IRS estimated that unreported income from legal sources grew at a 13 percent annual rate over roughly the last decade, from $93.9 billion in 1973 to $249.7 billion in 1981, while unreported income from the three selected illegal activities grew at a 17.7 percent annual rate, from $9.3 billion to $34.2 billion. To estimate unreported income from legal sources, the IRS drew mainly upon data on individual taxpayers from its Taxpayer Compliance Measurement Program, which audits a sample of income tax returns, and upon data from its Information Returns Program, which uses information from the payers of income. It developed estimates of unreported income from legal sources for individuals not filing returns by cross-checking information from two nationwide household surveys against its own records and those of the Social Security Administration. Finally, the IRS estimated unreported income obtained in the selected illegal activities from survey data and arrest records.

This approach to estimating the size of the underground economy has been subject to criticism. Some contend that the estimates derived from administrative records and surveys are likely to understate actual unreported income. They believe that estimates derived from monetary statistics offer a better gauge of underground activity and unreported income.

Aside from issues such as the underpayment of tax liabilities, the existence of an underground economy that may be growing relative to the recorded economy creates problems for analyses of public policy issues, including monetary policy. For example, policies developed from data on the recorded economy may not necessarily stabilize the total economy; or, movements in monetary aggregates that reflect changes in the underground economy may be interpreted as signaling change in the recorded economy. Thus policymakers need to assess the scope of the underground economy to see whether these potential issues deserve more explicit consideration.

This article evaluates estimates of the size and

growth of underground activity based on several monetary-statistic methods. The article also examines some of the reasons for the growth of per capita currency holdings, particularly in larger denominations—another phenomenon cited as evidence of underground activity.[4]

Currency-Ratio Method

The earliest monetary-statistic approach to estimating the size of the underground economy relies on an analysis of movements in the ratio of currency to checkable deposits—the currency ratio. In this technique the underlying assumption is that the currency ratio in the aboveground economy is constant over time. Because of this assumption an increase in the amount of money held as currency relative to that held in checkable deposits is interpreted as a relative rise in underground economic activity.[5] To implement the method, a benchmark period is selected that is assumed to be free of underground activities. "Normal" or aboveground currency in any period is then defined to be in the same proportion to actual checkable deposits in that period as total currency was to checkable deposits in the benchmark period; accordingly, underground currency is the difference between currency in circulation and estimated aboveground currency. The estimated size of the underground economy is determined as the product of underground currency and the income velocity (the ratio of income to money) of aboveground M1, which is the sum of aboveground currency and all checkable deposits. The last step in the calculation is based on the assumption that income velocity is the same in the underground and the aboveground sectors.

Currency-ratio estimates of underground gross national product appear in table 1, column 1. These figures suggest that the dollar level of underground activity was little changed until the middle 1970s, but almost tripled between 1975 and 1982, reaching $450 billion. As a percent of recorded GNP, the size of the underground economy remained roughly constant until the 1970s.

1. Computed underground GNP, alternative methods and selected years, 1950–82[1]

Year	Simple currency-ratio method	Modified currency-ratio method[2]	Tanzi's model of the ratio of currency to M2		Transaction-ratio method	
			(*TW*)	(*T*)	1939 base	1964 base[2]
	(1)	(2)	(3)	(4)	(5)	(6)
	Billions of dollars					
1950	15.9	21.5	14.5	9.4	27.6	43.1
1955	14.7	15.6	12.8	10.9	1.7	21.6
1960	17.3	17.1	20.7	13.2	−3.4	21.5
1965	31.6	38.6	26.3	17.1	9.6	44.3
1970	62.4	88.5	45.6	25.3	101.0	155.2
1975	150.8	246.0	77.0	46.6	467.3	567.1
1978	266.1	460.2	114.2	80.9	551.1	685.5
1979	317.8	558.5	130.7	88.6	628.4	779.2
1980	372.8	666.9	159.9	116.9	1,095.6	1,280.1
1981	427.1	767.6	n.a.	n.a.	1,765.6	1,999.2
1982	449.7	810.5	n.a.	n.a.	n.a.	n.a.
	Ratio to recorded GNP, percent					
1950	5.6	7.5	5.1	3.3	9.6	15.1
1955	3.7	3.9	3.2	2.7	.4	5.4
1960	3.4	3.4	4.1	2.6	−.7	4.2
1965	4.6	5.6	3.8	2.5	1.4	6.4
1970	6.3	8.9	4.6	2.6	10.2	15.6
1975	9.7	15.9	5.0	3.0	30.2	36.6
1978	12.3	21.3	5.3	3.7	25.5	31.7
1979	13.1	23.1	5.4	3.7	26.0	32.2
1980	14.2	25.3	6.1	4.4	41.6	48.6
1981	14.5	26.0	n.a.	n.a.	59.8	67.7
1982	14.6	26.4	n.a.	n.a.	n.a.	n.a.

1. For a description of each method see the text.
2. It is assumed that underground GNP equals 5 percent of observed GNP in 1964.

n.a. Not available.

The proportion then increased sharply after 1975, to a sizable 14.6 percent in 1982.

Modified Currency-Ratio Method

In 1980 Feige modified the currency-ratio method to make it conform more closely to what he believed were the actual practices in the underground economy.[6] Whereas the simple currency-ratio method postulates that currency is the exclusive medium of exchange in the underground economy, Feige argues that some firms and households use checks for such transactions because they perceive that the ease of doing so outweighs the costs of leaving a "paper" audit trail. He also contends that the underground sector is service-oriented. Because fewer intermediate transactions occur in the production of services, the amount of money balances per dollar of output is smaller in this sector than in the aboveground sector. Feige therefore assumes that the currency ratio in the underground sector equals two and that the income velocity of underground money is 10 percent higher than its aboveground counterpart.[7]

The modified currency-ratio estimates of underground GNP for selected years are shown in table 1, column 2. For the mid-1960s, this method gives higher estimates of underground GNP than does the simple currency-ratio method; beginning in the 1970s the gap between the two estimates widens greatly; and by 1982 the modified currency-ratio estimate of underground GNP, at 26.4 percent of aboveground GNP, is almost twice the estimate derived from the simpler approach.

A Variant of the Currency-Ratio Method: Tanzi's Method

Another variant of the currency-ratio method has been used by Vito Tanzi to estimate underground activity.[8] Tanzi develops an explicit empirical model of the ratio of currency to M2 that links the size of the underground economy to the incentive to evade taxes. Specifically, the demand for currency relative to M2 rises whenever real per capita income or the rate of interest on time deposits (which are included in M2) falls, or whenever the share of wages and salaries in national income or the level of taxes rises. The last variable reflects the presumed pecuniary advantage of engaging in underground activity to evade taxes, with a step-up in tax rates fostering a relative rise in underground activity and inducing an increase in desired currency holdings relative to other balances in M2.

To calculate the size of the underground economy, Tanzi estimates his model using annual data for the years 1930 to 1980.[9] Two simulations are then conducted. In the first, all explanatory variables take on their actual historical values to produce a predicted currency series that is consistent with the actual tax rates in each period. In the second simulation the tax rates are set equal to zero rather than their historical values. The difference between the two predicted amounts of currency is Tanzi's estimate of the amount of money in use in underground activities. As in the simple currency-ratio method, the income velocities of underground and aboveground money balances are assumed to be identical. Underground GNP is therefore the product of the estimated stock of underground currency and the income velocity of aboveground M1 balances.

Table 1, columns 3 and 4, presents the size of underground activity estimated with this model using two tax measures: *TW*, a weighted average tax rate on interest income, and *T*, the ratio of total net tax payments to adjusted gross income. Because both sets of estimates remain in a relatively narrow range around 5 percent of recorded GNP, they provide a striking contrast to the previous currency-ratio estimates. The figures indicate only a slight upward trend in the relative size of the underground economy; even for 1980 (the most recent year for which data are available), Tanzi estimates that underground GNP equaled only 6.1 percent (*TW*) or 4.4 percent (*T*) of aboveground GNP.

Transaction-Ratio Method

The ratio of total monetary transactions to gross national product is the main ingredient of the second basic approach to estimating underground activity, the transaction-ratio method developed by Feige.[10] Feige proposes that monetary transactions in underground activity will be

recorded in measures of total transactions but excluded from recorded income. Thus changes in the ratio of transactions to income will reflect changes in the underground economy. The key assumption underlying Feige's approach is that total transactions, the sum of debits to checkable deposits and the total dollar volume of currency transactions, are proportional to total economic activity ("total" here meaning the sum of aboveground and underground activity). Because total transactions include direct transfer payments, which exhibit a changing pattern over time, and purely financial transactions, which have increased dramatically in response to various financial innovations, Feige reformulates his original assumption in terms of the proportionality between a *net* transaction measure and total income. To derive a net transaction measure appropriate for estimating underground activity, he adjusts gross transactions by deducting several categories of major financial transactions and direct transfers.[11]

Given these adjustments, the calculation of underground GNP proceeds in much the same fashion as in the currency-ratio method: aboveground transactions are determined as the product of the ratio of transactions to GNP in the benchmark period (which is assumed to be free of underground activity) and recorded GNP. The excess of actual transactions over aboveground transactions defines the level of underground transactions for any given year. Underground income can then be inferred from the benchmark ratio of transactions to income.

Table 1 lists alternative transaction-ratio estimates of underground GNP. The estimates in column 5 are based on a 1939 benchmark period, while those in column 6 assume that underground GNP was 5 percent of recorded GNP in a 1964 base period. The transaction-ratio estimates of the size of underground activity are even larger than those estimated from the currency-ratio methods, rising from approximately 10 or 15 percent of reported GNP in 1970 to more than 60 percent by 1981.

A Critique of the Monetary-Statistic Methods

This section evaluates the assumptions, procedures, and estimated size of the underground economy for each of the various monetary-statistic methods just described. The effort here (and in the next section, which looks at currency data) is to explain the observed behavior of currency and transactions in traditional, aboveground terms, and thus avoid an underground explanation except as a last resort. The aboveground explanations are firmly rooted in economic theory and established empirical work, while, as will be shown, a number of the underground arguments bear only a tenuous relation to accepted theory and empirical practice.

The starting point is the observation that all of the methods except Tanzi's yield sharply increasing ratios of underground GNP to aboveground GNP since the late 1960s, particularly after 1975. Such a pattern implies a sharp increase in the *total* GNP velocity of M1, the ratio of the sum of aboveground and underground GNP to the level of M1. Table 2 displays the level and growth rates of total GNP velocity for three monetary-statistic methods for some of the years given in table 1.[12] As the table indicates, the velocities of total and recorded GNP grew on average at an annual rate of between 3.1 and 3.5 percent from 1950 to 1970. From 1975 onward, however, the estimated growth rates of total GNP velocity accelerate relative to those for recorded GNP velocity, which stays close to its long-run historical trend rate of change. For example, total GNP velocity for the transaction-ratio method using the 1939 base grows at an annual rate of 7.6 percent from 1975 to 1981, more than double the rate for the period 1950–70. Those who believe that both money demand and the aggregate economy are stable in the long run will regard such a sharp change in the trend of velocity as highly unlikely.

Another reason for skepticism stems from the apparent contradiction between such large estimates of underground activity and the results of a substantial body of empirical work. Although the underground economy may influence the relative amount of currency holdings, many other important factors are ignored by the advocates of the currency-ratio approach. The behavior of currency relative to checkable deposits or to M2 can, in fact, be explained with some accuracy by standard empirical demand equations that do not rely upon underground motives. Specifically, the standard macroeconomic approach to analyzing

2. Implied total income velocity of money by alternative methods of estimating underground activity, and recorded velocity[1]

Year or period	Simple currency-ratio method	Modified currency-ratio method	Transaction-ratio method		Velocity of M1 based on recorded GNP
			1939 base	1964 base	
	Implied velocity				
1960	3.705	3.704	3.559	3.735	3.583
1965	4.378	4.420	4.245	4.455	4.186
1970	4.996	5.120	5.180	5.436	4.701
1975	5.967	6.301	7.077	7.428	5.436
1978	6.929	7.483	7.742	8.125	6.168
1979	7.242	7.879	8.064	8.463	6.400
1980	7.487	8.220	9.288	9.748	6.558
1981	7.864	8.656	10.977	11.520	6.870
1982	7.691	8.479	n.a.	n.a.	6.711
	Average annual growth of implied velocity				
1950–70	3.1	3.4	3.3	3.1	3.5
1975–81	4.7	5.4	7.6	7.6	4.0
1975–82	3.7	4.3	n.a.	n.a.	3.1

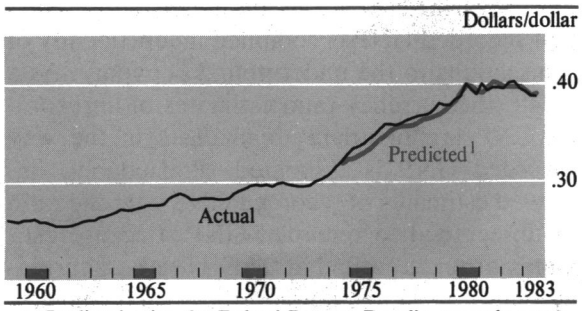
1. Velocity is measured as the ratio of the sum of aboveground (or recorded) GNP and underground GNP to an M1 measure.

n.a. Not available.

these ratios involves a model based on theories of the demand for money involving aboveground transactions or portfolio considerations.

As an indication of what the standard approach can explain, charts 1 and 2 display actual and predicted values of the alternative ratios from simulations using the Federal Reserve Board's quarterly econometric model.[13] The explanation rests primarily on interest rates, income, wealth, and prices, with no reference to underground activity. In general, the model's demand equations for the components of M1 and M2 fit the data fairly well. The equation explaining the demand deposit component of checkable deposits, however, includes a shift variable for the two and one-half years from 1974:3 to 1976:4; when this variable is removed, the model's equation, like most conventional demand equations, overpredicts demand deposits and, by implication, underpredicts the ratio of currency to checkable deposits.

1. Actual and predicted ratios of currency to transaction deposits

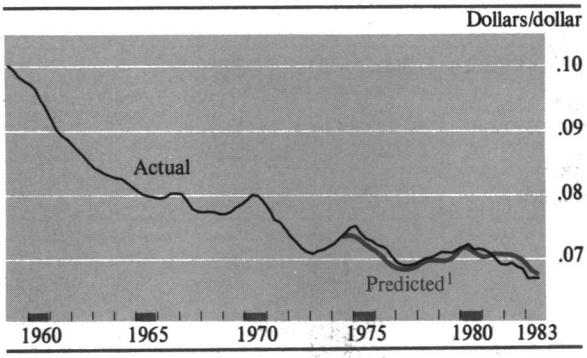


1. Predicted using the Federal Reserve Board's quarterly econometric model.

Although this failure to explain the spurt in the actual currency ratio might be viewed as evidence of an active underground economy, another explanation is perhaps more likely. The Board model and other models provide no evidence of any unexplained strength in currency itself during this period; the shortfall in predicting the currency ratio stems principally from the unexplained weakness in demand deposits.[14] Extensive analysis of this weakness in demand deposits suggests that, facing persistently high opportunity costs of holding demand deposits, deposit holders sought to improve their money management techniques.[15] This quest was aided by improvements in computer and telecommunications technology, by the development of various cash management procedures such as cash concentration accounts and remote disbursement facilities, and by the growing use of new financial instruments that complemented many of these new techniques.

More important, the simple and modified cur-

2. Actual and predicted ratios of currency to M2



1. Predicted using the Federal Reserve Board's quarterly econometric model.

rency-ratio methods ignore these ongoing technological and financial innovations because they both assume that the ratio of checkable deposits to currency is constant in the two sectors. This assumption is made solely for technical convenience, of course, but it does have the effect of denying any role whatsoever to important economic determinants of these ratios such as interest rates. For example, the introduction of negotiable order of withdrawal accounts nationwide in 1981 and Super NOW accounts in 1983 lowered the opportunity cost of holding transaction accounts (the difference between an open market yield such as a Treasury bill rate and the own yield on the NOW account), making it relatively more attractive to hold balances in such accounts rather than in currency. Because the currency-ratio methods do not account for such aboveground innovations, they incorrectly attribute the induced change in the observed currency ratio to developments in the underground economy.

While conventional empirical work predicts the ratio of currency to M2 fairly accurately, Tanzi's model does also, so that his work merits a closer look.[16] In contrast to standard money demand approaches, which assume only motives related to aboveground transactions and portfolio considerations for holding currency and deposits, Tanzi's approach includes an explicit tax term in the demand equation for currency relative to M2 to represent the incentive to evade taxes. The quality of the resulting estimates of the size of underground activity depends on the accuracy of the underlying specification and estimation of the tax effect. The data reveal that the positive relationship between the ratio of currency to M2 and taxes is strong *only* for the period from 1930 to 1945.[17] Indeed, the relationship breaks down in the postwar period, and thus Tanzi's model provides little evidence that an increase in taxes spurs an increase in underground activity.[18]

Each of the three currency-based methods involves arbitrary choices about relative income velocities, the proportions of currency and checkable deposits used in the two sectors, and the benchmark period.[19] In addition, these methods contain the implicit assumption that recorded GNP covers no underground activity. In fact, the Bureau of Economic Analysis (BEA) of the Department of Commerce compiles estimates of the national income and product accounts in recognition of the many distortions in the underlying sources of GNP data created by legal activity in the underground economy. Although BEA's success in limiting such distortions may be debated, it is erroneous to assume that reported GNP reflects only the aboveground economy. For instance, underreporting of income for tax purposes creates few serious statistical problems in the national income accounts because IRS data do not play an important role in developing estimates of national income. But where IRS sources must be used, reported income is adjusted on the basis of the IRS audit studies. In general, BEA prefers methods that impute a value of income, and such methods often are independent of whether a recorded monetary transaction has taken place. As a result, recorded GNP reflects at least some part of the legal underground economy. Furthermore, recently BEA has sought to adopt procedures that better estimate the component of underground activity that is conceptually consistent with its measures of income and product. The currency-ratio methods, nevertheless, are based on the assumption that recorded GNP is compiled independently of transactions in the underground economy. As a result, the currency-ratio estimates of unrecorded GNP are invariant to changes in the way recorded GNP is estimated. Presumably, improved estimates of recorded GNP alter the ratio of unrecorded to recorded GNP. Because estimates of underground activity based on currency-ratio procedures do not reflect such changes, those estimates are probably overstated.

The transaction-ratio method is more difficult to evaluate because, unlike the demand for currency and checkable deposits, total transactions are not subject to any established theory. Casual inspection of the ratio of transactions to income suggests that it often moves positively with interest rates. In a recent paper, Porter and Offenbacher offer a partial explanation for such movements based on an inventory model of money holdings under uncertainty.[20] This paper shows that the volume of debits to demand deposits for business firms should be positively related to both interest rates and a scale variable (which serves as a proxy for the size of the firm) and negatively related to the costs of transactions.[21] With this model, several of the major movements in the ratio of transactions to GNP can be explained without reference to factors associated with the underground economy. Nonetheless, additional theoretical and empirical work is required before the Porter–Offenbacher results can be viewed as firmly established.[22]

In comparison to the various currency-ratio methods, the transaction-ratio method has several distinct advantages, at least in principle. The method makes no assumption regarding the relative income velocities in the aboveground and underground sectors. It also treats currency and deposits in a symmetric fashion; that is, the method does not assume that currency is the exclusive medium of exchange in the underground sector or that currency and deposits are used in fixed ratios in each of the two sectors. Moreover, improved estimates of recorded GNP appropriately modify the resulting estimate of the ratio of underground GNP to recorded GNP; for example, an increase in recorded GNP will necessarily lower this ratio.

On the other hand, the transaction-ratio method requires the specification of a "benchmark" transaction ratio in the aboveground sector; as with the other methods, the choice of this ratio is a critical assumption. In practice, however, data limitations are the single most important problem in implementing the transaction-ratio method: the dollar volume of many significant financial transactions is simply not compiled either privately or publicly.[23] For example, direct measurements of the turnover of the currency stock do not exist, and indirect procedures must be used to estimate it.

The recent estimates of underground GNP from the transaction-ratio method suggest that increases in the transaction ratio itself are attributable largely to transactions in checkable deposits, not currency.[24] Because the likelihood of "catching" a participant in an underground transaction is probably higher when checkable deposits rather than currency are used, it seems counterintuitive to associate all of the implied increase in total income arising from the increase in checkable deposits with underground transactions. In addition, the 18.1 percent annual rate of growth in total income velocity in 1981 is about four times recorded velocity growth for that year (see table 2). Such a large increase in velocity is also unlikely and suggests that some purely financial component of total transactions has not been properly netted out, so that an upward bias has been imparted to the estimated transaction ratio for that year. Similar surges in velocity growth during other recent periods may also be due to various netting-out problems arising in the compilation of total transactions.

Thus far, this article has evaluated several methods that rely on an analysis of monetary statistics to estimate underground economic activity in the United States: the simple and modified currency-ratio methods, Tanzi's variant of the currency-ratio approach, and the transaction-ratio method. According to all of these methods, the relative size of the underground economy has increased over the last decade; Tanzi's estimates of underground GNP are relatively small (about 5 percent of recorded GNP), while those produced by the transaction-ratio method exceed 60 percent of recorded GNP. Unfortunately, each of these methods has significant problems of a methodological nature or in data requirements that call into question the basic reliability of the approach.

An Evaluation of the Currency Data

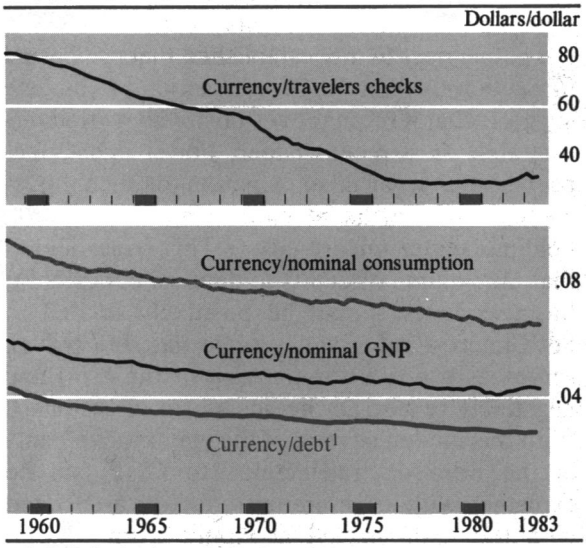
Although the monetary-statistic methods described earlier for estimating underground activity are not based exclusively on currency data, many observers believe that the most compelling evidence concerning the scope of the underground economy may be inferred from such data. They point to the remarkably high level of cur-

rency holdings per household and the sizable proportion that is held in large denominations. At the end of December 1983, currency holdings stood at almost $1,970 per household in the United States; just under 40 percent of this stock, or nearly $800, was in hundred-dollar bills. Even allowing for the currency that is held by businesses in cash registers and by financial institutions as vault cash or that has been lost or destroyed, these magnitudes seem to contradict everyday experience.[25] Even if a substantial fraction of the currency stock were held abroad, the implied level of domestic currency holdings would still be strikingly large.

It is difficult to account for such currency holdings in terms of a transaction theory of the demand for money. As a rough calculation, suppose that all income were received in the form of currency and all households were paid biweekly. The average household would then receive about $1,060 every two weeks.[26] If, in addition, all of the currency were spent on goods and services during the two-week interval between income payments, the typical household would on average have about half its original pay, or about $530, in the form of currency. The substantial discrepancy between this predicted amount and actual currency balances, which are roughly four times as large, indicates the nature of the difficulty for a transaction-based model of currency. Other factors, however, may account for holdings greater than the predicted $530. For example, many households are paid less frequently than biweekly, at least for part of their income, and some households may hold currency for precautionary reasons and as a store of wealth. On the other hand, several factors work to reduce currency holdings below this hypothetical average. Many households are paid exclusively by check, and many use checkable deposits for a substantial part of their transactions. In addition, households that are adding to their wealth by saving or are paid more frequently than biweekly will hold less currency. On balance, it seems difficult to explain the actual level of currency holdings solely on the basis of aboveground transactions; an underground explanation for these levels must be taken seriously.[27]

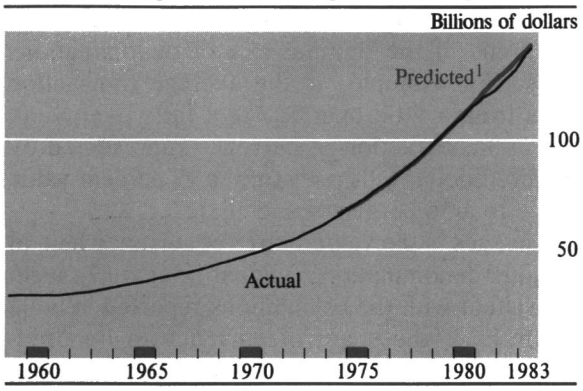
Despite the high and somewhat puzzling *level* of currency balances per household, the evidence does not suggest that *growth* in currency has been excessive relative to deposits or expenditures. Charts 2 and 3 show that, on balance over the past 20 years, total currency has been declining, not rising, relative to other financial aggregates such as M2, travelers checks, or domestic nonfinancial sector debt—or relative to nominal expenditures such as GNP and measured personal consumption. In the case of M2, this movement is not surprising because the average nominal rate of return on the noncurrency part of this aggregate has moved up sharply over this period as a result of deregulation and higher nominal interest rates, while the nominal pecuniary return on currency remained at zero. A similar declining pattern is apparent, at least through the mid-1970s, for the ratio of currency to travelers checks. This decline is somewhat unexpected because travelers checks, like currency, bear no nominal rate of return but, unlike currency, leave a paper trail. Thus, if underground activity were relatively more important over this period, that ratio should have risen, other things equal. Finally, currency movements over the past years have been highly predictable in conventional empirical models of money demand, which relate real currency holdings per capita to real consumption expenditures per capita and the opportunity cost of holding money but which make no reference to the underground economy (chart 4).

3. Ratios of currency to travelers checks, consumption, GNP, and debt



1. Debt in the domestic nonfinancial sector.

4. Actual and predicted holdings of currency



1. Predicted using the Federal Reserve Board's quarterly econometric model.

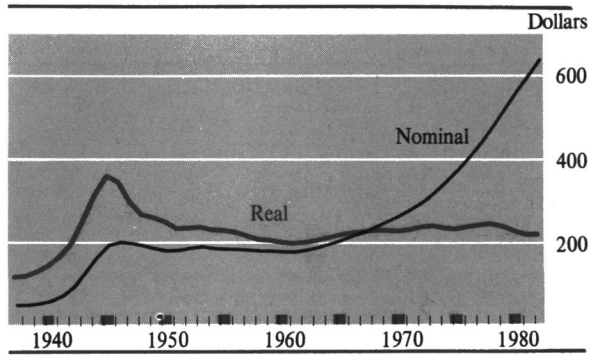
The accurate prediction of the growth of currency balances by conventional empirical models may be fortuitous, of course. Because currency holdings are the sum of aboveground and underground holdings, a relative decline in currency holdings in the aboveground sector owing to changes in payment practices may offset a relative increase in underground currency holdings, thereby leaving the total unaffected. For example, aboveground currency holders may have economized on currency by using credit cards more frequently. By itself, however, this factor seems unlikely to provide the full explanation because credit cards account for only a small proportion of estimated total currency transactions—just over 2 percent in 1981.[28] In addition, use of currency in the aboveground economy may have declined because a growing fraction of individuals has been paid by check rather than with currency. This possibility has not been explicitly recognized in the standard currency demand relationship; however, the predictions of the currency equation in the Board's quarterly model are not materially altered when it is accounted for, as Tanzi did, by using the ratio of compensation of employees to national income as an explanatory variable.

Still another development calls into question the view that currency holdings provide evidence of a growing underground economy. Since the mid-1950s, aggregate currency balances (including vault cash) have only about kept pace with inflation so that real currency holdings per capita have changed only slightly (chart 5). If real per capita holdings instead of total currency holdings were used in the monetary-statistic approach, the relative size of the underground economy would be approximately the same over most of the postwar period.[29]

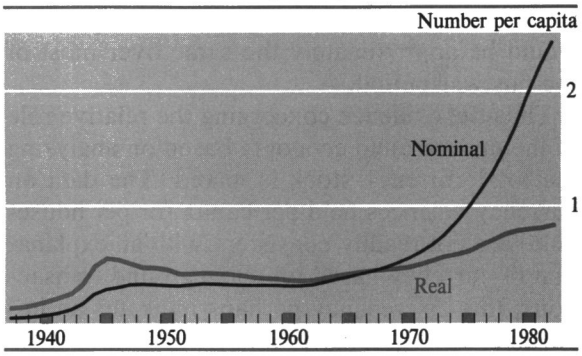
Thus the evidence concerning the relative role of the underground economy based on analyzing the total currency stock is mixed. The data on currency balances held per capita (or per household) are not readily consistent with an explanation of currency based on aboveground transactions; this discrepancy perhaps indicates an important underground presence. On the other hand, currency movements over time appear to be explained reasonably well by ongoing developments in the aboveground sector.

Although the historical data on aggregate currency do not provide unequivocal support for a growing underground economy, proponents of that view often point to the rising proportion of hundred-dollar bills in the currency stock. They contend that most large aboveground transactions are paid for by check, and that the growing use of large-denomination bills must be attributed principally to a growing volume of underground transactions. Per capita holdings of hundred-dollar bills rose from about 0.5 in 1966 to about 2.4 in 1982 (see chart 6). Even in real terms (1967 dollars), the change in per capita holdings of hundred-dollar bills is substantial: from 0.5 in 1966 to about 0.8 in 1982 (chart 6). Does this relative shift to large-denomination bills mask increased underground economic activity, or does it reflect the responses of aboveground transactors to changes in the economic environment? With regard to the latter possibility, it should be noted that, since 1969, the hundred-dollar bill has been the largest currency denomination issued.[30] Thus increases in the price level

5. Currency balances per capita



6. Holdings of $100 bills per capita



that tend to increase the dollar size of transactions should, other things equal, spur the use of hundred-dollar bills relative to other denominations because they are more convenient in large-scale transactions.[31]

The importance of hundred-dollar bills in the mix of denominations can be evaluated with a model recently proposed by J. S. Cramer.[32] Cramer assumed that transactors attempt to economize on the number of physical units of currency used in an exchange of a given transaction size. Table 3 presents the results of applying Cramer's model to the various bill denominations in the United States for various ranges of transaction size.[33] The estimates were constructed under the assumption that all transactions up to a certain size were equally likely to occur while, beyond that size, the likelihood of a transaction declined as its size increased. For example, transactions of $2,000 were assumed to occur less frequently than transactions of $1,000.[34] Table 3 presents calculations from this

3. Value shares of bills of various denominations in optimal mix of denominations for selected average transaction sizes

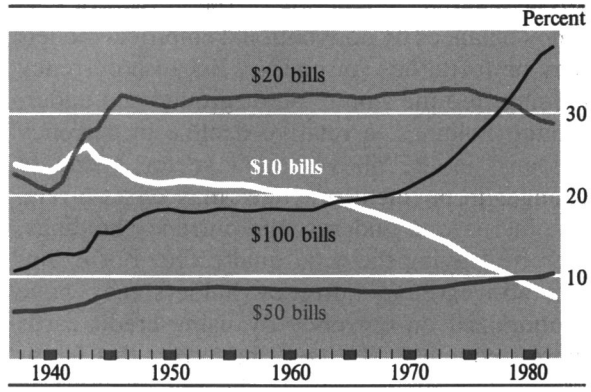
Percent

Average transaction (dollars)	$100 bills	$50 bills	$20 bills
12.69	12	13	22
25.38	19	20	33
38.08	25	28	29
50.77	31	32	23
63.46	37	34	18
76.15	43	32	14
88.85	49	29	13
101.54	56	25	12
114.23	62	21	10
126.92	66	19	9

model indicating that, as the dollar size of individual transactions increases, the proportion of hundreds in the optimal mix of denominations rises. For example, as the average transaction goes from a little over $25 to a little over $100, the optimal fraction of currency represented by hundred-dollar bills rises from a 19 percent value share to a 56 percent value share.

Changes in the value shares of currency held in various denominations, shown in chart 7, seem consistent with the calculations reported in table 3. In 1978, the share of currency in hundred-dollar bills surpassed the share in twenty-dollar

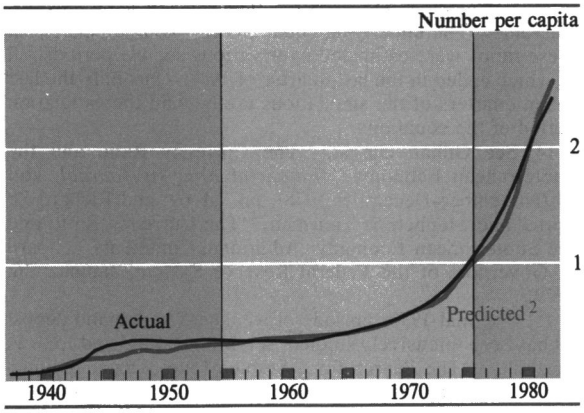
7. Value shares of currency held in various denominations of bills



bills. The chart shows that the last time a similar event occurred was in 1942, when the amount of money represented by the twenty-dollar denomination became larger than the amount held in ten-dollar bills.[35] Over the period from 1942 to 1978, consumption expenditures per capita grew from $657 to $6,049. Thus, assuming that total transactions per capita and the average size of transactions moved together, there is an above-ground explanation for the increasing share of hundred-dollar bills: per capita consumption expenditures were more than nine times as large, while the size of the denomination in which the largest proportion of currency was outstanding was five times as large.

Another, related explanation that focuses on the use of hundred-dollar bills in the above-ground sector has been developed. Essentially, this explanation describes the relationship between per capita holdings of hundred-dollar bills and the price level.[36] The predictions from the

8. Holdings of $100 bills per capita, actual and predicted [1]



1. In-sample years are not shaded. Out-of-sample years are.
2. Predicted by regression equation described in Richard D. Porter and Amanda S. Bayer, "Evaluating Underground Economic Activity in the United States Using Monetary Statistics," Staff Study (Board of Governors of the Federal Reserve System, forthcoming), appendix C.

implied empirical equation are shown in chart 8. The equation performs quite adequately in the out-of-sample period, explaining a substantial part of the recent increase of per capita holdings of hundred-dollar bills.

These theoretical and empirical results suggest that the expansion in the use of hundred-dollar bills is related principally to normal economic and institutional forces at work in the aboveground economy. While the amount and form of currency holdings appear suspiciously large, the interaction between increases in the price level and the size pattern of available currency denominations appears to account for the actual mix of denominations in currency holdings.

Summary and Conclusion

This article has examined several estimates of the size and growth of underground activity that have been developed using monetary statistics. Nearly all of these estimates imply an expansion in the proportion of underground activity relative to total activity and a large rise in the total income velocity of money since 1970. Both currency-ratio methods utilize readily available data, but they depend on several questionable assumptions. The most critical are (1) a constant ratio of currency to checkable deposits in the aboveground sector despite changes in important economic determinants such as interest rates and the own rate of return on negotiable order of withdrawal and automatic transfer service accounts; (2) an erroneous belief that recorded gross national product is estimated with no recognition of legal underground activities; and (3) either no use of checkable deposits or the fixed proportional use of currency and checkable deposits in the underground sector. Although the transaction-ratio method avoids these pitfalls, it has severe data limitations, relating especially to the separation of purely financial transactions from others.

Evidence has also been gleaned from an explicit model of the ratio of currency to M2, which relates the size of the underground economy to the incentive to evade taxes. In contrast to the other estimates, this method suggests that the relation of the underground sector to total economic activity has not changed significantly. This method also makes several questionable assumptions, however: (1) the ratio of underground GNP to recorded GNP does not vary with the method for compiling recorded GNP; and (2) underground transactions involve only currency. Moreover, the method fails to find evidence of the predicted tax effect when estimation is restricted to the postwar period.

Although the enormous size of currency holdings per capita or per household is puzzling, it can be explained by standard demand relationships that relate currency holdings per capita to real consumption expenditures per capita and to the opportunity cost of holding currency. Increases in the price level combined with explicit recognition of the available denominations of currency appear to account for changes in the mix of currency denominations.

The analysis of underground activity has not progressed enough to permit a reliable estimate of the scope of such activity from an analysis of monetary data. Given current techniques, these data do not convincingly support the hypothesis that the share of the underground economy in the total U.S. economy has grown recently. Perhaps as more satisfactory data and techniques emerge, better estimates can be developed. In any event, the issues raised in attempting to measure underground activity by these methods pose some challenging questions regarding the use of currency and deposits as transaction media in the aggregate economy. □

FOOTNOTES

1. Peter M. Gutmann, "The Subterranean Economy," *Financial Analysts Journal,* vol. 33 (November–December 1977), pp. 26–27; Edgar Feige, "How Big Is the Irregular Economy?" *Challenge*, vol. 22 (November–December 1979), pp. 5–13.

2. *Estimates of Income Unreported on Individual Tax Returns,* U.S. Department of the Treasury, Publication 1104 (9–79), p. ii.

3. The estimates of income from illegal sources are preliminary; see Internal Revenue Service, Assistant Commissioner for Planning, Finance, and Research, *Income Tax Compliance Research* (U.S. Department of the Treasury, July 1983), pp. 9, 39.

4. Richard D. Porter and Amanda S. Bayer, "Evaluating Underground Economic Activity in the United States Using Monetary Statistics," Staff Study (Board of Governors of the Federal Reserve System, forthcoming), provides a more detailed and more technical discussion of the issues examined in this article.

5. The method was originally suggested by Phillip Cagan to evaluate the upward movements in the currency ratio in World War II; see Phillip Cagan, "The Demand for Currency Relative to the Total Money Supply," *Journal of Political Economy*, vol. 66 (August 1958), pp. 303–28. The method was later adopted by Peter Gutmann, "Subterranean Economy;" and it was subsequently modified by Edgar Feige, "A New Perspective on Macroeconomic Phenomena: The Theory and Measurement of the Unobserved Sector in the United States Economy—Causes, Consequences, and Implications," paper presented at the 1980 meetings of the American Economic Association.

The initial estimates of underground GNP made by both Gutmann and Feige covered a period when the amount of deposits in other checkable accounts such as ATS, NOW, and Super NOW accounts was small; those investigators thus ignored these accounts in their work and used the ratio of currency to demand deposits. In the last few years these new accounts have grown rapidly and have tended to substitute for demand deposits rather than for currency; as a consequence, the ratio of currency to demand deposits has risen for reasons totally unrelated to underground activity. Thus, in this article, the currency-ratio estimates are based on the ratio of currency to checkable deposits. As a reference point, appendix table A.1 presents estimates of underground activity using the ratio of currency to demand deposits.

6. Ibid.

7. Ibid, pp. 19–22.

8. Vito Tanzi, "The Underground Economy in the United States: Annual Estimates, 1930–80," International Monetary Fund, *Staff Papers*, vol. 30 (June 1983), pp. 283–305.

9. See Porter and Bayer, "Evaluating Underground Activity," for a more detailed discussion of the estimates.

10. Feige, "How Big?" and "New Perspective."

11. See Porter and Bayer, "Evaluating Underground Activity," for a detailed discussion of the data used and steps involved in compiling the adjusted series on transactions.

12. Appendix table A.2 presents currency-ratio and modified currency-ratio velocity measures for the narrower definition of the currency ratio, the ratio of currency to demand deposits.

13. The charts represent dynamic simulations of the equations starting in the third quarter of 1974 and extending through the third quarter of 1983. (Appendix B of Porter and Bayer, "Evaluating Underground Activity," presents the equations as well as a brief explanation of their structure.) In these simulations the determinants of the ratios—interest rates, real income, and so forth—take on their actual historical values. The underlying equations for the components of these ratios were estimated over various sample periods, all of which ended in the last quarter of 1981. Thus only the last seven quarters of the simulations are beyond the estimation period of the equations.

14. See Gillian Garcia, "The Currency Ratio and the Subterranean Economy," *Financial Analysts Journal,* vol. 34 (November–December 1978), pp. 64–69; and Richard D. Porter and Stephen S. Thurman, "The Currency Ratio and the Subterranean Economy: Additional Comments" (Board of Governors of the Federal Reserve System, January 26, 1979).

15. The mid-1970s episode of weakness in demand deposits has been intensively studied; see John P. Judd and John F. Scadding, "The Search for a Stable Money Demand Function: A Survey of the Post-1973 Literature," *Journal of Economic Literature*, vol. 20 (September 1982), pp. 993–1023; Richard D. Porter, Thomas D. Simpson, and Eileen Mauskopf, "Financial Innovations and the Monetary Aggregates," *Brookings Papers on Economic Activity, 1:1979*, pp. 213–29; Thomas D. Simpson and Richard D. Porter, "Some Issues Involving the Definition and Interpretation of the Monetary Aggregates," in *Controlling the Monetary Aggregates III*, Federal Reserve Bank of Boston Conference Series No. 22 (October 1980), pp. 161–234; and Jared Enzler, Lewis Johnson, and John Paulus, "Some Problems of Money Demand," *Brookings Papers on Economic Activity, 1:1976*, pp. 261–80.

16. One difference between the conventional models and Tanzi's model is that the latter uses the old definition of M2, which includes only M2 deposits held at commercial banks.

17. Even for the period before 1946, the specification can be questioned because it does not take into account the introduction of deposit insurance, which altered the demand for currency relative to M2.

18. When the estimation period for Tanzi's model is restricted to the postwar years 1946–80, the estimated coefficient for the tax variable has the wrong sign when *T*, the ratio of total tax payments to income, is used; that is, as taxes increase the ratio of currency to M2 falls. With *TW*, the weighted average tax rate on interest income, the estimated tax coefficient does not come close to being statistically significant. See Porter and Bayer, "Evaluating Underground Activity," appendix B-10.

19. In Tanzi's method, the benchmark assumption concerns the threshold level for taxes. Tanzi assumes that underground activity develops as soon as any tax is placed on output. However, this threshold tax effect could conceivably be triggered at some value above zero.

20. Richard D. Porter and Edward K. Offenbacher, "Financial Innovations and Measurement of Monetary Aggregates" (Federal Reserve Bank of St. Louis, forthcoming).

21. The particular proxy used for transaction costs is described in Simpson and Porter, "Some Issues," table 4, form number 1, p. 283. Also, for simplicity the scale variable is taken to be recorded GNP.

22. The simulation results from the Porter–Offenbacher model are merely within-sample predictions and thus are not particularly strong evidence regarding the explanatory power of these equations. See Porter and Bayer, "Evaluating Underground Activity," for a more detailed discussion of the results.

23. Ibid.

24. Ibid.

25. At the end of December, vault cash was about 12 percent of the total. There are no available data indicating total currency held by businesses. Robert D. Laurent has estimated that lost currency has never accounted for more than 4 percent of currency in circulation; see his "Currency in Circulation and the Real Value of Notes," *Journal of Money, Credit, and Banking*, vol. 6 (May 1974), pp. 213–26.

26. This estimate assumes $2.4 trillion in aggregate annual disposable income and 87.3 million households in the United States.

27. Per household or per capita figures may be misleading and may not indicate the median level of currency balances. For example, in 1975 currency holdings per capita were about $330. This figure may seem high for that time, but it does not necessarily imply that a person chosen at random would hold such an amount; some would hold more and some less. A relatively small fraction of the population might well hold a sizable portion of the total stock of currency. Such a distribution would be consistent with the size distribution of demand deposit holdings, which is highly skewed: in 1975, 0.6 percent of demand deposit account holders held about half of all the demand deposits, according to estimates by the Federal Deposit Insurance Corporation. If the size distribution of currency were the same, it would imply that, excluding the 0.6 percent of the population that held the largest amounts, currency holdings per capita in 1975 would be only $165, or half of overall per capita holdings.

28. See Porter and Bayer, "Evaluating Underground Activity."

29. Because the total economy has grown over this period, the relative constancy of real currency holdings per capita implies, other things equal, that the underground economy has shrunk relative to the aboveground economy. In terms of the Board's estimated currency equation, the increase in the opportunity cost of holding currency and autonomous improvements in managing currency apparently have offset the increased level of real transactions, thereby leaving real currency holdings per capita about unchanged.

30. Denominations larger than $100—$500, $1,000, $5,000, and $10,000 bills—have not been printed since 1946. They have not been issued since 1969.

31. The importance of fifty-dollar bills should increase somewhat also, but hundred-dollar bills appear, as will be shown below, to be more efficient over a wider range of transaction sizes.

32. J. S. Cramer, "Currency by Denomination," *Economics Letters*, vol. 12 (1983), pp. 299–303.

33. We are indebted to Gary Anderson of the Board staff for his technical assistance in compiling this table.

34. Formally, the size distribution of transactions is assumed to be uniform (all transactions are equally likely) up to a given point and to follow Pareto distribution beyond that point; that is, the distribution function for transactions was specified to be

$$f(x) = \begin{cases} c \text{ if } x \leq \beta \\ c \left(\frac{\beta}{x}\right)^{\alpha+1} \text{ if } x \geq \beta, \end{cases}$$

where β is the upper limit of the uniform portion of the distribution and $c = \alpha/\beta(\alpha + 1)$. The parameter in the Pareto distribution α was set equal to 1.65. This is the approximate value estimated for a variant of this model discussed below to explain per capita holdings of hundred-dollar bills. See Porter and Bayer, "Evaluating Underground Activity," appendix C.

35. A comparison of table 3 and chart 6 for fifty-dollar bills, however, raises one problem with this explanation. The table suggests that fifty-dollar bills should have surpassed twenty-dollar bills before they were overtaken by hundred-dollar bills, but the chart indicates that that event never occurred at all. If the analysis used to explain hundred-dollar bills in the text is basically correct, the reconciliation of the share data for fifty-dollar bills must require a different Pareto parameter estimate or different size distribution for transactions than that set out in note 34. An econometric investigation of these questions is currently being conducted by members of the Board's staff.

36. Basically, the regression model discussed in the text is derived from the following assumptions: (1) the size distribution of transactions is a Pareto distribution for transactions above a given size; (2) in response to inflation, the size distribution shifts in proportion to the change in the price level; and (3) hundred-dollar bills are used in large transactions. See Porter and Bayer, "Evaluating Underground Activity," especially appendix C, for further discussion of this model.

Appendix tables appear on the following page.

A.1. Computed underground GNP using the ratio of currency to demand deposits[1]

Year	Simple currency-ratio method	Modified currency-ratio method
	Billions of dollars	
1950	15.9	21.4
1955	14.7	15.5
1960	17.3	17.0
1965	31.7	38.6
1970	62.6	88.6
1975	152.2	248.3
1978	280.2	489.9
1979	359.6	649.8
1980	445.2	829.2
1981	683.5	1375.2
1982	832.1	1748.4
	Ratio to recorded GNP, percent	
1950	5.6	7.5
1955	3.7	3.9
1960	3.4	3.4
1965	4.6	5.6
1970	6.3	8.9
1975	9.8	16.0
1978	12.9	22.6
1979	14.9	26.9
1980	16.9	31.5
1981	23.1	46.6
1982	27.1	56.9

1. The estimates of underground GNP in this table are derived via the simple and modified currency-ratio methods, as described in the text, but use the ratio of currency to demand deposits as opposed to the ratio of currency to total checkable deposits.

A.2. Implied total income velocity of money using the ratio of currency to demand deposits to estimate underground activity[1]

Year or period	Simple currency-ratio method	Modified currency-ratio method
	Implied velocity	
1960	3.706	3.704
1965	4.381	4.423
1970	5.001	5.124
1975	5.986	6.324
1978	7.076	7.684
1979	7.645	8.444
1980	8.108	9.120
1981	9.983	11.882
1982	10.620	13.111
	Average annual rate of growth of implied velocity	
1950–70	3.1	3.2
1975–81	8.9	11.1
1975–82	8.5	11.0

1. Velocity is measured as the ratio of the sum of aboveground (or recorded) GNP and underground GNP to an M1 measure.