ESTIMATING THE VALUE OF UNRECORDED ECONOMIC ACTIVITY IN SOUTH AFRICA

B Kantor*

Abstract

This paper attempts to measure the value of unrecorded economic activity in South Africa. The value of unrecorded activity is inferred from the quantity of notes in circulation. It is observed that the ratio of notes in circulation to official measures of gross domestic product and expenditure has increased significantly in recent years. This development has coincided with greater freedoms for black South Africans to migrate to the urban areas of South Africa and to engage in economic activity there. Differences between the actual demand for notes and predictions of such demands, derived from models of the demand for notes, as estimated by regression analysis for the period 1950 to 1980, are used to infer the volume of unrecorded activity. Annual data is applied.

The models used to estimate the demand for notes are a mixture of uni-variate time series models, combined with bank debits, interest rates and inflation as other independent variables. The time series models provide very good fits. The addition of the other independent variables add little explanatory power.

Consideration is also given to the reference of possible changes in the demand for notes by the banking system.

Both on theoretical and empirical grounds it is shown that much of the increase in the demand for notes unrecorded outside rather than inside the banking system.

1. Introduction

Casual observation would suggest that the informal sector in South Africa has become more important. Trading activity in the streets of the cities and townships has become pervasive. Informal markets seem very active and many more people in and outside the townships seem to be involved in selling, providing services and manufacturing products.

There are a number of possible reasons for such activity. Of primary importance has surely been the greater freedom exercised by black South Africans over recent years to migrate to and trade in the established urban areas. Over the same period, the formal

*Professor in Economics, School of Economics, University of Cape Town, Private Bag, Rondebosch 7700, Republic of South Africa.

sector itself has not provided many additional jobs[1]. Relatively slow growth of formal activity and, perhaps more important, moral suasion to pay what are regarded as 'living' wages, as well as a more closely regulated formal labour market, have encouraged the formal sector to employ fewer, better-paid workers (President's Council, 1987). In addition, increases in income tax rates and sales tax rates have no doubt encouraged informal, that is to say unrecorded for tax purposes, activity. (For an account of changes in tax rates, see RSA (1988: 29)).

Unreported or unrecorded income or output or expenditure has always had to be inferred for the purposes of keeping national income accounts. Among the more important of the items imputed for national income accounting purposes are depreciation allowances and rental income from owner-occupied homes (Canada, 1975). Income from self-subsistence farming should also be imputed for the purposes of measuring national income. A general movement of workers from, perhaps, underrecorded subsistence activities to more market-related activity could, therefore, be an important influence on measured GDP, assuming again that these market-related activities were properly recorded.

2. Applying the residual

National income accounting methods are surely most reliable when income is earned and expenditure undertaken in formal markets and where market participants keep good and honest records for their own purposes and for the purposes of meeting their tax obligations. Of course, the opportunities and the incentive to avoid taxation by under-reporting income are obvious enough. The owner-manager of a small business has perhaps the advantage over the manager in larger corporations in the opportunities to under-report income or to "live off" the business. It is because of the incentive to under-report, or not to record, income which may be taxed that national income accountants regard estimates of expenditure as more reliable than estimates of income and output (Macafee, 1980). In Britian, for example, estimates of final expenditure consistently exceed estimates of income or output. By definition, if properly measured, the value of final expenditure should equal the value of incomes earned and of value added. In practice, the different sides of the national accounts will never balance exactly. Furthermore, if the patterns of spending should alter as a result of changes in the underlying economic structure, the sample surveys used for estimating expenditure would need to be updated.

In South Africa, however, unlike Britain, measures of output and income are regarded as intrinsically superior to measures of final expenditure. Thus, when estimates of expenditure exceed the, somewhat independently, obtained estimates of output and income, estimates of expenditure are reduced accordingly to match up with output. Such a procedure would be perfectly acceptable, if the residual applied to expenditure estimates had an expected value of zero. That is to say, the chances of expenditure

[1]Between 1980 and 1987 recorded employment grew by 5% and the economically active population of South Africa, excluding that of the Republics of Transkei, Bophuthatswana, Venda and Ciskei, grew by 19%. *Source:* *Quarterly Bulletin of Statistics*, CSS, South Africa, 1988; *S.A.R.B. Quarterly Bulletin*, March 1988, S.104.

exceeding or falling short of output were about the same. Recently this has not been the case. Recorded expenditure has consistently exceeded recorded output and the residual item represents an important proportion of total expenditure and output (see Figure 1). In previous attempts to measure the size of the informal sector, using quarterly national income data, similar observations were made about the relative significance of expenditure and output measures and the biased nature of the expenditure residual (Kantor, 1987 and 1988). These comments attracted the following responses from the head of the Economics Department at the SA Reserve Bank.

> "The widening discrepancy between the output measure and the expenditure measure which assumed unprecedented proportions during recent years when estimated aggregate expenditure consistently exceeded estimated aggregate output, is caused to some extent, as you rightly suggest in your paper, by the growth in the informal sector. It can be argued that estimates of private consumption expenditure, at least, capture some informal sector activity whereas output estimates fail to do so.
>
> Bench-mark estimates of private consumption expenditure are calculated *inter alia* from the results of the "Household Income and Expenditure Pattern" surveys. As you may be aware, these surveys are periodically conducted among White households by the Central Statistical Service and among Black, Asian and Coloured households in selected urban areas by the Bureau of Market Research of the University of South Africa. Since these household surveys concentrate on the spending behaviour of households irrespective of the source of their income, it can safely be assumed that spending out of income earned through "informal" or untaxed economic activity would be duly reported in the relevant questionnaires. Information obtained from these sources is supplemented with information on income and population distribution as obtained from the periodic population censuses conducted by Central Statistical Service as well as from various ad hoc studies on consumer expenditure and household income. Total private comsumption expenditure estimates as compiled by the Reserve Bank for bench-mark years can therefore be said to account for the spending of income earned in the informal sector. (Revised private consumption expenditure bench-mark estimates for 1985 are to be compiled in the near future from the recently published reports on household income and the expenditure and population census for that year). The output measure of the GDP does not at present adequately include informal sector activities. They are normally based on censuses of business concerns with the necessary legal approval to do business. Labour remuneration and the gross operating surpluses, which represent the major elements of the output estimates, are extracted from the financial results contained in these census reports. Unlicensed business undertakings and other self-employed people without locus standi in the business world and fearing possible prosecution will probably refrain from rendering returns in terms of the various official censuses. Output estimates based on sectoral census results will thus contain an element of undercoverage.
>
> The long delay between the taking of censuses and the finalisation of census results is an additional source of error in the output measures of gross domestic

production. For example, the most recent results for a census of manufacturing relate to 1978.

It is doubtfull whether income reported in household surveys will be helpful in assessing the size of income generated in the informal sector. Households could be expected to be rather reluctant in reporting informal sector income, for fear of it being brought into the tax ambit".

3. The demand for money and the size of the informal sector

This particular study, in part, replicates the earlier papers, replacing quarterly with annual data, using the methods developed first by Feige for the analysis of unrecorded activity. The approach adopted is to infer unrecorded activity by observation of the behaviour of the monetary system and in particular the demand for currency relative to bank deposits and more convenient media of exchange (Feige, 1979, 1980, 1982 and 1984). The 'unexplained' demand for narrow money or currency is regarded as a proxy for unrecorded economic activity.

It may be presumed that informal, unrecorded activity, will be transacted largely through the exchange of notes and coin rather than by banks. Informal sector participants either find the use of banking accounts either too costly or too dangerous, in that they may attract the attention of the tax authorities.

A characteristic feature of economic development is an increase, over time, in the degree of financial sophistication. Using the banking system has obvious advantages over shipping large quantities of notes to settle indebtedness. Thus lower transactions costs and growing financial sophistication may be expected to reduce the demand for notes and increase in the demand for alternative media of exchange. The relationship between transactions costs and demands for notes can be fully explored within the so-called inventory approach to money demand (Baumol, 1952 and Tobin, 1956).

The inventory model attempts to derive the optimum stock of money held for transaction purposes, given the costs of holding money and the costs of switching between other assets and money. The costs incurred in holding money are firstly the interest income foregone and secondly the costs of switching, which include the time incurred in so doing. These transaction costs are assumed to be fixed independently of the amount of the transaction. Thus the average fixed cost of holding money declines when the amount of cash increases, whereas the interest cost of holding money of course increases with the amount held. There is some optimal period between transactions where total costs of holding money are minimised. On the further assumption that sums cashed in are spent evenly over the period, the formula for the optimum average money balances held is as follows:

$$M/P = \sqrt{c(F/P)/2R}$$

where M is the nominal amount of cash held, P the price level, R the nominal market rate of interest and F the fixed transaction cost and c the constant rate at which the money holder runs down money holders or spenders. The result is known as the square

root formula because it relates the demand for money to the square root of the rate of spending c. (For a modern derivation of this result, see Barro (1987)).

Thus it may be seen that, by applying calculus, an increase in the fixed costs of switching will cause the demand for real cash to increase while an increase in interest rates would cause the demand for real cash to decline. Any reduction in transactions costs, for example, through the availability of machines from which cash may be drawn (automatic tellers) by reducing the fixed transactions costs, will increase the frequency with which the machines will be used to obtain cash, so reducing the optimum period between transactions and reducing the average amount of real cash held. It should be understood that the machines save the user the time previously spent in the banking hall. Time spent paying in cash has been reduced, as salaries and wage payments are increasingly made electronically by transfers from the bank account of the employer to that of the employee. This method for paying wages and salaries is, of course, one of the reasons why the demand for notes could be expected to decline over time.

Should economic agents reduce their demands for real cash in response to more convenient banking generally, then the banks themselves would be required to reduce their own demands for non-interest-bearing notes to meet the smaller, but more frequent withdrawals of their customers. Fewer notes in circulation outside of the banks, at any moment in time, would be associated, presumably, with a smaller inventory of notes inside the banks, unless there were some economies of scale in the banks' demand for notes, that is, if the optimum inventory of notes held by banks, to meet withdrawals, increased with the number of outlets for cash. Automatic teller machines, with their own cache of notes, are in effect the equivalent of additional tellers and additional branches. Also from the banks' perspectives, cash in the vaults or in the tills satisfy reserve requirements demanded by the Central Bank. Other reserves may also be non-interest-bearing, so reducing the incentive to minimise holdings of notes.

4. The evidence

In Figure 2, the South African ratio of notes to GDP at factor cost over the period 1923 to 1987 is indicated. As may be seen, the ratio of notes to GDP peaks during the Second World War, and declines more or less continuously to 1980, whereafter the ratio picks up again. The increase in the demand for notes during the war is surely ample testimony to the importance of illegal activity then. The increased real demand for notes during the 1930's probably had something to do with the mistrust of the banking system during the Depression. The reversal of this note to measured economic activity ratio recently is, of course, of particular interest. Statistics on the value of bank debits in South Africa are available from 1938. The ratio of bank debits to GDP or GDE or the real value of bank debits may be taken as a measure of financial sophistication, or of the use of the banking system. Bank debits have become increasingly important in South Africa and continued to become more important after 1980. (See Figures 3 and 4) Official statistics for GDE and GDP, at market prices in addition to GDP at factor costs are

only available after 1946[2].

The ratio of notes in circulation in South Africa to GDP at factor cost and at market prices and to GDE, all increased after 1980 despite growing financial market sophistication and deregulation, as indicated by the series for bank debits. (See Figure 5) Such a reversal of the long-term trend cannot be explained, it is argued, except in terms of a growing volume of informal, unrecorded activity.

5. Regression analysis

The first method used here to infer the volume of unrecorded activity is to estimate, by way of regression analysis, the note to activity ratios observed between 1950 and 1980 with the ratios predicted by such models are then used together with the actual note issue between 1980 and 1987, to estimate 'total' activity. Differences between such estimates of total activity and the official statistics are regarded, therefore, as a measure of unrecorded activity. It should be noted also that the relationship between the note issue and expenditure over the entire period under observation is a generally close and reliable one, as may be seen in Figure 6.

It is the recent differences between the trend in the note issue and official measures of expenditure that represent *prima facie* evidence of proportionately more unrecorded activity since 1980.

The second approach adopted was to estimate an equation linking economic activity, GDP or GDE as measured, to the note issue, as opposed to the ratio of notes to GDP or GDE and to use that equation, together with the actual note issue, to estimated 'true' activity after 1980. The difference between such estimates and officially recorded activity may then also be regarded as an estimate of unrecorded activity.

The econometric results generated by both methods are presented in Table 1. As may be seen, the statistical results obtained for all of these models are highly significant. The method partly employed was the construction of a univariate time series model for the estimation of the ratio of notes to different official measures of economic activity for the period 1950-1980. The within period fits are reported in Table 1 and may be seen to be highly significant. The time series models were then used to forecast the ratio out of period from 1981 to 1987. The forecast ratios replicate the previous structure over the forecast period. These forecast ratios were then used together with the actual note issue to provide an estimate of 'total' economic activity.

When economic activity was regressed directly on the note issue, the residuals in such simple models also proved highly correlated, and so it was necessary to model in the systematic time series behaviour of the residual. This is shown to be an autoregressive

[2]The sources for national income and monetary data have been various *Quarterly Bulletins of the South African Reserve Bank*, especially the earlier volumes and the *Supplement to the Quarterly Bulletin of September 1981*, a statistical presentation of South Africa's national accounts for the period 1946-1980. The source for GDP at factor cost before 1945 was *South African Statistics*, Department of Statistics, Pretoria, 1974.

process or order two[3].

As may also be seen, (Equations 8-10), the inclusion of bank debits as an explanatory variable did not add to the explanatory power of the models. The reason for this is probably because of the co-variance of real debits and real economic activity. Real rates, real activity and real debits rise together.

In Table 2, the actual and estimated ratios of notes to the different measures of economic activity are reported. In Table 3, the official measures of economic activity and those inferred from the predicted note ratios are presented. In Table 4, the ratios of the different measures of activity to their official equivalent are presented. In Table 5, the results obtained via the alternative method are presented.

The precise influence of interest rates on the demand for currency in South Africa is not identified by regression analysis. When the ratio of notes to GDE or GDP is regressed on the long term interest rates, the coefficient on interest rates appears as statistically significant. However a high degree of auto-correlation is also revealed. When the equations are adjusted for auto-correlation, the interest coefficient loses its statistical significance. In an equation which regresses the growth in real notes on the growth in real GDE and interest rates, the coefficient for the interest rate variable appears to be statistically insignificant.

When inflation is substituted for interest rates as an explanatory variable, similar results to those obtained with nominal interest rates are obtained (see Equation 17, Table 1, Equations 14, 15 and 16).

6. The demand for reserves by South African banks

This analysis is complicated to a degree by changing preferences for reserves revealed by the South African commercial banks. South African banks may keep reserves, in the form of cash in their vaults or deposits at the Reserve Bank. Vault cash may consist of notes or coin and coin may include gold coins in the form of Kruger Rands.

It may be seen in Figure 7 that the proportion of the commercial banks' cash reserves held in the form of deposits declined after 1986. The increase in reserves held at the Reserve Bank after 1980 and the subsequent decline after 1982 was the result of changes in reserve requirements. (It should be appreciated that the increase in the ratio of notes to GDE in South Africa precedes these developments.)

There are two possible explanations for the change in the composition of cash held by the banks. Firstly, as the demand for notes to use outside the banks has grown with increased activity in the informal sector, the banks have been required to hold additional notes as a working inventory. The other possible explanation for developments is that the banks are not only holding additional notes, but additional gold coins in the form of Kruger Rands, which also meet reserve requirements. In recent years, an active forward market in Kruger Rands has developed which enabled

[3]All regressions were performed using TSP Version 4.2.

the banks to earn a positive rate of return on their holdings through forward sales at a premium to the spot price. Thus, in effect, not all reserve holdings have been non-interest-bearing.,

Nevertheless, over the same period 1984-1988, notes in circulation have grown significantly faster than M3, the widest definition of money. The ratio of notes in circulation to M3 is illustrated in Figure 8. Slower growth in the banking sectors' liabilities would have implied slower growth in the cash reserves required of banks. If, at the same time, the demand for notes outside of the banking system were growing, this would have forced the banks to keep a larger working inventory of notes. This factor, in addition to the incentive to hold Kruger Rands, may be responsible for the decline in the banks' ratio of reserve balances at the Reserve Bank to other reserves held by banks, to be observed in Figure 7.

7. Summary

This study has attempted to infer unrecorded economic activity via the otherwise unexplained demand for notes in South Africa. The real demand for notes increased significantly after 1980. Depending on the method used for inferring unrecorded activity, GDP at market prices may be estimated as between 16 and as much as 41 per cent higher than official recorded levels in 1987.

Clearly there can be no absolutely accurate measure of what is unrecorded. The results reported here should be regarded as suggestive of the recent growth in unrecorded activity in South Africa. There are two elements of the analysis about which firm conclusions can be drawn. Firstly, that the estimated residual should be added to GDP rather than subtracted from GDE and secondly, that the recent growth in demand for currency indicates an important new element of unrecorded economic activity and structural change in South African economy.

References

Barro, R J (1987): *Microeconomics*, 2nd edition, New York: John Wiley & Sons.

Baumol, W (1952): "The Transactions Demand for Cash: An Inventory Approach", *Quarterly Journal of Economics*, 66.

Canada (1975): "A Guide to the National Income and Expenditure Accounts", *National Income and Expenditure Accounts*, Volume 3, Statistics Canada.

Feige, E L (1979): "How Big is the Irregular Economy?" *Challenge*, 21, 5-12

Feige, E L (1980): "A New Perspective on Macro Economic Phenomena: The Theory and Measurement of the Unobserved Sector of the United States Economy", *American Economics Association Meetings*, Denver, Colorado, August.

Feige, E L (1981): "The U.K.'s Unobserved Economy: A Preliminary Assessment", *Journal of Economic Affairs*, 1(4).

Feige, E L (1980): *Observer-Subject Feedback: The Dynamics of the Unobserved Economy*, Leiden: Leiden University Press.

Feige, E L (1984): "Microeconomics and the Unobserved Sector", Bloch, W and Walker, M (eds): *Taxation:*

An International Perspective, Proceedings of an International Symposium, Vancouver: The Fraser Institute.

Kantor, B S (1987): "How big is the Informal Sector? Implications for the Share Market", *PL News,* November, 1-3.

Kantor B S (1988): "Economic and Political Freedom in South Africa: The Role of the Informal Sector", *Southern African Freedom Review,* 1(4), 29-39.

Macafee, K (1980): "A Glimpse of the Hidden Economy in the National Accounts", *CSO Economic Trends,* Central Statistical Office, No. 316.

President's Council: (1987): *Report of the Committee for Economic Affairs on a Strategy for Employment Creation and Labour Intensive Development,* Cape Town: Government Printer.

RSA (1989): *Statistical/Economic Review Budget 1988/89. Indicators of Tax Revenue.* Cape Town: Government Printer.

Tobin, J (1956): "The Interest-Elasticity of Transactions Demand for Cash", *Review of Economics and Statistics,* August, 38, 241-247.