THE ECONOMIC RECORD, VOL. 81, NO. 255, DECEMBER, 2005, 394–403

Australia's Cash Economy: Are the Estimates Credible?*

TREVOR BREUSCH

Centre for Social Research, Research School of Social Sciences, and School of Economics, Faculty of Economics and Commerce, The Australian National University, Canberra, Australia

The method of 'excess sensitivity' of Bajada (1999, 2000, 2001) indicates a large underground economy in Australia, with estimates of unrecorded income approximately 15 per cent of official gross domestic product. These estimates concern policy-makers, especially those agencies responsible for national accounts, tax collection, economic stabilisation and law enforcement. We show that the method exhibits a severe form of non-robustness, in which the results change markedly with a simple change in the units of measurement of the variables. There is a separate problem in which a key parameter is set to an unrealistic value that makes the estimates many times too high.

I Introduction

Policy-makers and citizens in Australia are alarmed by research findings that unreported cash transactions might be funding an underground economy as large as 15 per cent of official gross domestic product (GDP). The figure comes from a series of works by Christopher Bajada, including an article in the *Economic Record* (1999), a report for the Australian Tax Research Foundation (2001) and a research monograph (2002). All three of these works use the same method and contain the same core analytical material, and report the same 15 per cent result, although the focus and the time periods differ slightly. Another recent article in the *Economic Record* (Bajada, 2003), takes the earlier results as given and explores the business cycle properties of the underground economy. These research findings and their implications for social and economic policy have been given considerable attention in newspapers and the electronic media.[1]

The reasons for concern about unobserved economic activity will differ between agencies, depending on their areas of responsibility and the nature of the activity. The Australian Bureau of Statistics (ABS), which is responsible for compiling national income accounts, believes it has computed GDP to within 1 or 2 per cent of accuracy (ABS, 2004). Against this confidence, 15 per cent of unrecorded income would be a considerable embarrassment. However, if the underground economy arises from activity that is itself illegal, it does not count as GDP in the ABS definition. But then the Australian Taxation Office is charged with taxing incomes and the sales of goods and services, regardless of whether the underlying activity is legal. Unreported income to the extent of 15 per cent of GDP

*I thank the Associate Editor and the referees for their advice on an earlier version of the present paper. Several colleagues also made helpful comments on a draft, and I have benefited from background discussions with the Australian Bureau of Statistics, National Accounts Branch.

JEL classification: E26, E41, E51, E62.

Correspondence: Trevor Breusch, School of Economics, Faculty of Economics and Commerce, The Australian National University, Canberra, ACT 0200, Australia. Email: trevor.breusch@anu.edu.au

[1] 'Underground economy' is the preferred term for those production activities that are themselves legal, but are concealed from the authorities for the purpose of evading regulation or taxation (OECD, 2002). Other colourful terms, such as 'black' or 'shadow' economy, are sometimes used with various shades of meaning.

represents a large slice of lost tax revenue. To political leaders and citizens alike, it is unfair if some people are evading taxation and regulation, at the expense of others who are complying with the law. Enforcement authorities at state and federal levels may see unreported income as a marker of criminal behaviour that needs to be investigated. Apart from the revenue and distributional implications of large-scale tax evasion, the Australian Treasury and Reserve Bank of Australia are concerned about misjudgement of economic policy if they rely on inaccurate national income data, particularly, if the amount of unrecorded activity is changing over time.

Many approaches have been used to estimate the extent of underground economic activity around the world; see Schneider and Enste (2000) or Bajada (2002, Ch. 3) for surveys of this literature. One general class of methods is called the 'currency demand' method because it compares various macroeconomic aggregates to see if the amount of currency (i.e. cash – notes and coins) in the hands of the public (i.e. outside the banking system) depends on factors that suggest unreported income. Studies of the demand for money show that holdings in aggregate will depend on many factors, including the volume of transactions that has to be financed, the availability of other payment methods such as electronic transfers of bank balances, and the cost of interest income forgone by holding cash. After these factors are taken into account, there may be patterns over time in cash use that cannot be accounted for by the economic transactions that are observed by the authorities. According to the currency demand method this extra cash use constitutes evidence of unreported income.

In particular, it has been suggested that unobserved cash transactions might respond to incentives such as high income tax rates and high levels of welfare benefits. Thus, Bajada uses the term 'excess sensitivity' to describe his particular version of the currency demand method. The points of departure in this approach are the way the response of public cash holdings to income tax rates and welfare payments is measured, and the steps that lead from a measure of sensitivity to an estimate of the relative size of the underground economy. However, we show that this method is unsound because it is sensitive to the units of measurement of the variables. A simple change in the units of the tax rate variable – from a percentage to a decimal fraction – will produce a vastly different inference about the size of the underground economy. We also consider modifying the method so that it is robust to such changes in the units of measurement, but find that the underground economy all but disappears in the robust procedure.

Even if we leave aside the problems in calculation of 'excess sensitivity', there is a crucial assumption in this work that is superficially attractive but utterly implausible. It is assumed that the amount of income in a year that is supported by a given stock of available cash is the same in the underground economy (where cash is the only form of money) as it is in the observed economy (where cash is but a very small part of the money supply). We explore the consequences of this assumption and find that, on this account alone, the estimated size of the underground economy in Australia is several times too high to be credible.

II Modelling Currency Holdings

The analysis of 'excess sensitivity' starts with an equation to represent the aggregate holding of currency in the economy, which is assumed to be homogeneous in prices and population[2]

$$C = f(Yd, R, \pi, E, Tr), \qquad \text{(B-1)}$$

where C is real currency per capita, Yd real disposable income per capita, R the interest rate, π the inflation rate, E private consumption expenditure as a percentage of GDP, and Tr a technological trend. By definition of disposable income, $Yd = Y - Tx + Wf$, where Y is national income (the same as GDP in this setting), Tx is taxes and Wf is welfare benefits (presumably both real, per capita). The model is then expanded so the tax and welfare variables are allowed a direct influence on currency holdings in addition to their impact through disposable income:

$$C = f(Y - Tx + Wf, R, \pi, E, Tx, Wf, Tr). \qquad \text{(B-2)}$$

The term 'excess sensitivity' refers to the direct effects of Tx and Wf on C, at given levels of Yd and the other variables in the equation.

The estimation model for currency demand is written as linear in logs of real per capita variables and some ratios. These new ratio variables are T for the average income tax rate, constructed as a percentage of GDP, and W for welfare benefits as a percentage of disposable income. The model is estimated as an error correction regression, with one lag and one difference on all variables (except the welfare variable, which has only the first difference[3]). Thus, the empirical model takes the form

$$\Delta \ln C_t = \boldsymbol{X}_t\beta + \boldsymbol{Z}_t\gamma + \varepsilon_t. \qquad (3)$$

[2] Equation numbers shown as (B-n) match the numbering in Bajada (1999), although similar equations can be readily identified with different numbering in the other Bajada works.

[3] The absence of the lag of the welfare variable is not explained.

Here, $\boldsymbol{X}_t$ is a vector of 14 or 15 elements, namely, the first difference and first lag of $\ln Yd_t$, $\ln R_t$, $\ln \pi_t$ and $\ln E_t$. There is also a linear trend, three seasonal dummy variables, the lag of $\ln C_t$ and an intercept.[4, 5] Similarly, $\boldsymbol{Z}_t$ is a vector of three elements, namely, $\Delta \ln T_t$, $\ln T_{t-1}$ and $\Delta \ln W_t$. The reason for the partition into two subsets of explanatory variables, $\boldsymbol{X}_t$ and $\boldsymbol{Z}_t$, will be seen shortly. This model is estimated by least squares. In Bajada (1999) the data are quarterly, not seasonally adjusted, and cover the period from June 1966 to June 1996.

III Excess Sensitivity

The expositions in Bajada (1999, 2001, 2002) describe the extraction of 'excess sensitivity' from a rearrangement of the estimation model. First, the model is rewritten with the dependent variable being nominal currency holdings, $C^* = P \times N \times C$, where P is the price level (taken to be the implicit GDP deflator), and N is population. Then some variables are shifted to the right-hand side of equation 3, by use of the relationship, $\Delta \ln C_t^* = \Delta \ln C_t + \Delta \ln P_t + \Delta \ln N_t$, where also $\Delta \ln C_t^* = \ln C_t^* - \ln C_{t-1}^*$. This results in the rearrangement of equation 3 as

$$\ln C_t^* = \tilde{\boldsymbol{X}}_t \tilde{\beta} + \boldsymbol{Z}_t \gamma + \varepsilon_t, \tag{4}$$

where $\tilde{\boldsymbol{X}}_t$ is an expanded vector that now includes $\ln C_{t-1}^*$, $\Delta \ln P_t$ and $\Delta \ln N_t$, and where $\tilde{\beta}$ has been expanded to match.

'Excess sensitivity' refers to the additional effect on nominal currency holdings C_t^* that is attributable to the tax and welfare variables, which in our notation are contained in the vector $\boldsymbol{Z}_t$. Bajada follows Tanzi (1982, 1983) in calculating the difference between two predictions of currency holdings. First, a prediction is made from equation 4 of total currency given the data in $\tilde{\boldsymbol{X}}_t$ and $\boldsymbol{Z}_t$. The formula for this prediction is obtained by dropping the error term and exponentiating both sides to yield[6]

$$C_t^* = \exp\{\tilde{\boldsymbol{X}}_t \tilde{\beta} + \boldsymbol{Z}_t \gamma\}. \tag{5}$$

Then a second prediction is made after dropping the term $\boldsymbol{Z}_t \alpha$, which contains the tax and welfare variables. The result is called the 'natural' or 'legal' level of currency, and is written as[7]

$$C_{wt}^* = \exp\{\tilde{\boldsymbol{X}}_t \tilde{\beta}\}. \tag{6}$$

The difference between total currency in equation 5 and 'legal' currency in equation 6 is called underground or 'illegal' currency, and is

$$C_{ut}^* = \exp\{\tilde{\boldsymbol{X}}_t \tilde{\beta}\} \times (\exp\{\boldsymbol{Z}_t \gamma\} - 1), \tag{7}$$

where $C_{ut}^* = C_t^* - C_{wt}^*$ is a different symbol from the original, but preferred because it is more effective as a mnemonic.

IV Underground Income

The connection between illegal currency and unobserved income is made through an assumption about the velocity of circulation, which is the ratio of the flow of income in a year to the stock of money. Velocity in Bajada is measured for observed net national income (NNI), which is represented by Y^*. The reason for using NNI instead of GDP here is the reasonable one that 'both consumption of fixed capital and net income paid overseas are most likely to involve very small amounts of cash' (Bajada, 1999, footnote 23, p. 337).[8] The velocity of observed NNI in legal currency is $V_t = Y_t^*/C_{wt}^*$.

Now comes the key assumption. Velocity of income in the underground economy relative to illegal currency is assumed to be equal to velocity of income in the observed economy relative to legal currency, that is,

$$V_t = \frac{Y_{ut}^*}{C_{ut}^*} = \frac{Y_t^*}{C_{wt}^*}, \tag{8}$$

where we use the symbol Y_{ut}^* for underground income to match C_{ut}^* for underground currency.[9] Rearranging equation 8 gives an expression for underground income, as follows:

[4] Equation 3 is (B-4) rewritten in our condensed notation. In Bajada (2001) and (2002), there is an additional dummy variable to indicate the introduction of goods and services tax (GST) in July 2000. In Bajada (2001) the tax variable is split into tax on households and tax on business.

[5] This representation of the interest rate R and the inflation rate π in logarithmic form is a questionable practice, because these variables usually arise as discount rates in the exponents of multiplicative models and, hence, go into their natural form when the model is linearised. The log transformation on the inflation rate can be problematic with recent data, because the quarter-on-quarter changes in the implicit GDP deflator are sometimes negative.

[6] Equation 5 is (B-5) rewritten in our condensed notation. Although it may be intuitive, the form of predictor is questionable. Under standard assumptions on the error term, this formula with the least-squares estimates replacing the parameters is not the best predictor, nor even an unbiased predictor, in the log-linear regression model. See Verbeek (2000, p. 49).

[7] Equation 6 is (B-6) rewritten in our condensed notation.

[8] NNI in Australia is, on average, just over 80 per cent of GDP, varying in the annual data between 79 and 84 per cent over the period used for estimation.

[9] Equation 8 summarises (B-8) and the text immediately following that equation. Observe what might be considered an inconsistency in notation: observed currency C_t^* is total and, hence, includes the underground component C_{ut}^*, while observed income Y_t^* excludes underground income Y_{ut}^*.

$$Y_{ut}^* = Y_t^* \times \left(\frac{C_{ut}^*}{C_{wt}^*} \right). \tag{9}$$

Then, with substitution from equations 6 and 7, the underground economy as a proportion of observed GDP is evaluated as

$$U_t = \frac{Y_{ut}^*}{Y_t} = \frac{Y_t^*}{Y_t} \times \frac{C_{ut}^*}{C_{wt}^*}$$
$$= \frac{Y_t^*}{Y_t} \times (\exp\{\mathbf{Z}_t \gamma\} - 1). \tag{10}$$

Equation 10 is not needed to implement the method (and not reported as an equation in the original studies), because the estimation can be achieved by numerically manipulating the two series of predicted currency holdings, together with the series for recorded income, as in equation 9. But equation 10 is useful for our purpose, because it allows the original results to be replicated without revisiting the regression analysis in any detail. All that are needed to evaluate the estimate in this form are a few variables from the national accounts and the γ coefficients from the estimated regression model.

V *Replicating the Calculation*

When the regression estimates reported in Bajada (1999, table 3) are used to evaluate equation 10, the resulting percentage of GDP in underground income is calculated as

$$\hat{U}_t = \left(\frac{NNI_t}{GDP_t} \right) \times (\exp\{0.061 \ln T_{t-1} + 0.183 \Delta \ln T_t - 0.049 \Delta \ln W_t\} - 1) \times 100. \tag{11}$$

This formula can be evaluated using the data that are described in Appendix I. The results are shown in Figure 1 as two different plots. Direct calculation of equation 11 gives the extremely choppy plot, which clearly contains a large seasonal component – and a lot of irregular variation as well. Thus, we see why a moving-average smoother is applied in the original studies 'to reduce the volatile short-term fluctuations' (Bajada, 1999, footnote 25, p. 377). Applying a similar filter to our calculation gives the 'smoothed' plot in Figure 1.

The smoothed plot appears very close to the one presented in Bajada (1999, figure 2, p. 377). Ours is at the same level, approximately 15 per cent of recorded GDP for most of the period; it has the same shape and timing of its cycles everywhere; and even in its fine detail it is nearly identical to the original.[10, 11]

We can gain further insight into this method of estimating an underground economy by simplifying equation 11. The last term inside the braces, $-0.049\Delta \ln W_t$, contributes less than 0.4 per cent to the value that is obtained for the underground economy, as measured by the average relative error when the term is dropped. If a smoothed time series was calculated without this term and plotted onto Figure 1, it would be difficult to distinguish the new line from the previous one. The second last term, $0.183\Delta \ln T_t$, contributes more than that to the overall calculation, but dropping the two terms together makes only 1.3 per cent error. Put differently, the simplification

$$\hat{U}_t \approx \left(\frac{NNI_t}{GDP_t} \right) \times (\exp\{0.061 \ln T_{t-1}\} - 1) \times 100, \tag{12}$$

captures 98.7 per cent of the original inference. This expression can be rewritten using the properties of logarithms as

$$\hat{U}_t \approx K_t \times \left(T_{t-1}^{0.061} - 1 \right), \tag{13}$$

where K_t is NNI_t expressed as a percentage of GDP_t, a value that varies annually between 79 and 84 per cent for the whole period. In a slightly different form, using the approximation $e^x - 1 \approx x$ for small x, equation 13 may also be written approximately as

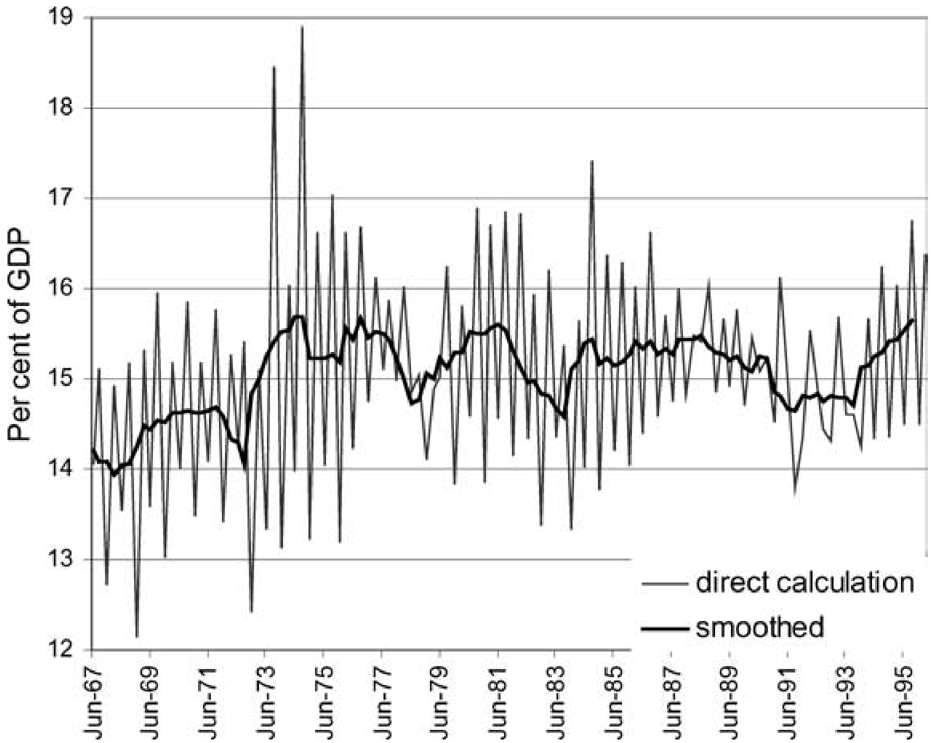
$$\hat{U}_t \approx K_t \times 0.061 \ln T_{t-1}. \tag{14}$$

Thus, from equation 13 or 14, apart from the almost constant 20 per cent discount in moving from *GDP* to *NNI*, with a small amount of variation in that discount factor from year to year, the principal driver of the inference about the underground economy by the method of 'excess sensitivity' is the income tax

[10] The strong seasonality in the unfiltered series occurs despite some seasonal smoothing that has already been applied to the tax series on which the variable T is based. See Appendix I for a description of the data. Some readers may regard this residual seasonality, which is large relative to the phenomenon being measured, as a warning sign that the method is too reliant on irrelevant features of the data. Alternatively, it might be argued (as by a referee) that it simply points to a less than satisfactory treatment of seasonality in the modelling.

[11] The filter is described as a 'four-period centred moving average' (Bajada, 1999, footnote 25, p. 377), but it appears that a four-period simple moving average is actually used. The centred form (which covers five time points in a weighted pattern) induces too much smoothness, and does not match the original plot as well as the simple filter.

FIGURE 1
Alleged Cash Economy in Australia (Per Cent of GDP; Source: See Section V and Appendix I)



rate. As income tax as a percentage of GDP goes up and down, so will the inference about the underground economy. The two variables are tied together in a close mathematical relationship.

VI Sensitivity to Units

One criterion of good analysis is that its findings are *robust*, which means that the results are not too much dependent on small changes in the assumptions that are central to the analysis, and are invariant to even large changes in any conditions that are peripheral to the system being analysed. In particular, relationships of substance are equally true or false in whatever units of measurements are chosen to describe them. The same substantive results must be obtained, for example, whether income is measured in thousands or millions of dollars, and whether tax rates are measured as percentages or decimal fractions.

Consider again Bajada's model and suppose now that taxes are measured in different units, not as a percentage of GDP, but as a fraction of GDP. So instead of being numbers like 15 or 20 per cent, the tax rate is now a decimal like 0.15 or 0.20. Imagine that we are totally thorough about this change of units, adopting it at the start of the empirical analysis and maintaining it consistently through to the end. Nothing is different in equations B-1 and B-2, because these are theory equations, and in these equations tax is a quantum amount not a proportion of income. The issue of units will have an impact on the estimation model, such as shown in equation 3, which can be expanded somewhat as

$$\Delta \ln C_t = \boldsymbol{X}_t\beta + \gamma_1 \ln T_{t-1} + \gamma_2 \Delta \ln T_t + \gamma_3 \Delta \ln W_t + \varepsilon_t. \quad (15)$$

If the new decimal variable is written as T_t^d to distinguish it from the old percentage variable T_t, the relationship is $100 \times T_t^d = T_t$. In equation 15, the change of units implies $\ln T_{t-1} = \ln 100 + \ln T_{t-1}^d$ and $\Delta \ln T_t = \Delta \ln T_t^d$. When the estimation model is written in the new units it becomes,

$$\Delta \ln C_t = \gamma_1 \ln 100 + \boldsymbol{X}_t\beta + \gamma_1 \ln T_{t-1}^d + \gamma_2 \Delta \ln T_t^d + \gamma_3 \Delta \ln W_t + \varepsilon_t, \quad (16)$$

where the only thing changed is the intercept in the model. As expected, when the model is re-estimated in the new measurement system, all the same coefficient estimates will be obtained, except the intercept will now be larger by an additional amount of $\gamma_1 \ln 100$ (approximately 4.61 times the slope coefficient on $\ln T_{t-1}$).

Suppose now these new estimates are used to extract a measure of 'excess sensitivity' in the same manner as before. The change in the measurement unit and its consequences can be traced through equations 4–7. The only alterations will be replacement of $\ln T_{t-1}$ by $\ln T^d_{t-1}$ in $\boldsymbol{Z}_t$ and the addition of an amount of $\gamma_1 \ln 100$ to the value of the intercept in $\tilde{\beta}$. In equation 10, all that is needed is the updated $\ln T^d_{t-1}$, because the intercept has cancelled from the expression for 'excess sensitivity'. Similarly, the new computational formula, equation 11, is the same as before, but with the new decimal measure $\ln T^d_{t-1}$ in $\boldsymbol{Z}_t$ in place of the old percentage measure $\ln T_{t-1}$.

But if this new version of equation 11 is computed and smoothed in the same manner as shown in Figure 1, the result is completely nonsensical. The values are all negative, in the range -8 to -11, and, hence, they make no sense at all when interpreted as the percentage of underground income in relation to GDP. We see that the calculation of the underground economy depends in an entirely arbitrary way on the units of measurement used for the tax ratio variable.

What has gone wrong? The problem comes from measuring 'excess sensitivity' as the effect of dropping the term $\boldsymbol{Z}_t\gamma$ in the predictor of currency holdings. Total currency in equation 5 is invariant to the change of units in T from percentages to proportions, because the replacement of $\ln T_{t-1}$ by $\ln T^d_{t-1}$ is exactly compensated by the addition of $\gamma_1 \ln 100$ to the value of the intercept in $\tilde{\beta}$. But 'legal' currency in equation 6 is not invariant to the change, because it carries the new intercept, but not the terms in T and W. The partition of total currency into 'legal' and 'illegal' components is, therefore, dependent on the units of measurement adopted for T. This shows how the substantive result on the size of the underground economy is sensitive to the measurement system.

The origin of the problem is the ad hoc conception of 'excess sensitivity' that is used. Simply dropping $\boldsymbol{Z}_t\gamma$ from the predictor does not correspond to any meaningful setting of a counterfactual against which the extra currency used in the underground economy can be measured. It is clearly not the same operation as setting the tax and welfare rates to zero, as used in many other methods of estimating the underground economy (see e.g. Tanzi, 1982), because the logarithm of the zero is not defined. Neither is it the same as the setting the counterfactual to be some minimum values of these variables, perhaps a conceptual minimum or the historical low, as adopted in some other studies (again, see Tanzi, 1982). Neither does it describe a counterfactual world in which tax and welfare rates have no effect on currency demand, because in that case the predictor would be re-estimated from a model that excludes those explanatory variables.[12]

VII Possible Solution

The problem we have isolated arises because a change of the units of measurement alters the intercept in the equation as well as the term that measures 'excess sensitivity'. This suggests a possible solution may be found in redefining the latter to include that part of the intercept that is affected. The intercept in a regression equation is nothing more than a linear combination of the means of the variables in the regression, with the regression slopes as the coefficients of the linear combination. This is so, both in a theoretical expression of the model in terms of population parameters, and as a relation between the estimated coefficients and the sample means of the variables. It is an elementary result that fitting an intercept in a regression model is exactly the same as estimating the model with no intercept, but with the variables transformed as deviations from their means. The intercept estimate can be recovered subsequently in a simple formula that relates the regression slopes and the sample means of the variables.[13]

The advantage of the deviation-from-means formulation in the present context is the 'excess sensitivity' term is then invariant to changes in the units of measurement. From equation 10, the calculation of the percentage of income in the underground economy will have the form

$$U_t = \frac{Y_t^*}{Y_t} \times (\exp\{\tilde{\boldsymbol{Z}}_t\gamma\} - 1), \qquad (17)$$

where now $\tilde{\boldsymbol{Z}}_t = \boldsymbol{Z}_t - \bar{\boldsymbol{Z}}$ contains the variables $\Delta \ln T_t$, $\ln T_{t-1}$ and $\Delta \ln W_t$ all transformed into deviations from their sample means. The estimate of γ will be the same as before, but now the change in units will be innocuous because the variables are in the

[12] Another indication of the ad hoc nature of 'excess sensitivity' is its dependence on only short-run or transient parameters in the regression model. Other approaches to measuring the underground economy from error-correction models emphasise long-run solutions and the dynamics of adjustment to equilibrium; see Hill and Kabir (2000) and Gadea and Serrano-Sanz (2002).

[13] For example, see Verbeek (2000, pp. 10–11).

form $\ln T_{t-1} - \overline{\ln T}$. Any change of the units of measurement of T is a multiplicative constant on T, which becomes an additive constant on both $\ln T$ and its average. When the variables are formulated into deviations-from-means, the constants arising from the change of units will cancel, leaving the 'excess sensitivity' term unaffected by the change.

It is, therefore, of some interest to evaluate the estimate of the underground economy in this modified form that is invariant to changes in units of measurement. Again this can be done from equation 10, using the same coefficient estimates, and with the same data for the variables apart from their transformation into deviations-from-means. The same smoothing is necessary to remove the high level of residual seasonality that was seen previously, and a similarly smooth result is obtained. The result is an estimate of the underground economy that varies between -1.5 per cent and $+0.8$ per cent of observed GDP. This estimate ranges from the illogical to the dubiously small.

Mathematically, this outcome is hardly a surprise. With the original definition of 'excess sensitivity', we noted earlier that the main driver of the inference about the underground economy is the variable $\ln T_{t-1}$. When T is measured as a percentage, a typical value is approximately $T = 15$ and, hence, the value of $\ln T$ is approximately 2.7; in this case, the inference is a large and positive underground economy. On the other hand, when T is measured as a decimal, a typical value is 0.15 and $\ln T$ is approximately -1.9; in this case, the underground economy is implausibly negative. The putative solution of mean correcting $\ln T$ ensures that this variable is approximately zero, certainly on average, so it is not surprising that the inference is an underground economy that is sometimes negative and at most very small.

VIII Key Assumption

Leaving aside the problems with the meaning and measurement of 'excess sensitivity', there is an assumption in this work that is key to the results, but so implausible that the estimates are incredible on that account alone. In moving from a measure of currency holdings to an estimate of underground income, a value is required for the velocity of circulation, that is, the ratio of the amount of income in a year to the stock of money. In equation 10, it is clear that the estimated size of the underground economy is directly proportional to the value of velocity in the hidden sector. Also, embodied in that calculation is the assumption that velocity in the underground sector is the same as the ratio of income to currency in the observed sector of the economy.

The assumption of equal velocities of currency in the two sectors is introduced on the grounds that 'There is very little we can do about such an assumption' (Bajada, 1999, footnote 24, p. 377). At least superficially, equal velocities seem a reasonable baseline assumption, a neutral one, and perhaps even a necessary one. However, demand for currency is not the same in the observed and underground sectors, because the role of cash is very different in the two sectors. In the underground economy, cash is the only form of money that is available for settling transactions without leaving an audit trail, whereas cash has a much smaller role to play in the observed economy. The two ratios of income to currency will be very different in the two sectors, and, hence, the velocity that results from this hypothesis is both extreme and untenable. The implied value of velocity is also much higher than used in other literature that uses currency demand models to estimate the underground economy. The resulting estimate of underground activity is inflated in direct proportion to the inflated value of velocity.

In the underground cash economy, where currency is the only medium of exchange available to those who wish to conceal their transactions, the ratio of income to currency is close to the textbook idea of velocity. It is, for that sector, the ratio of the total flow of income to the total stock of money available for the settling of transactions. Although the idea of velocity is clear enough in concept here, the issue in the underground economy is that neither the numerator nor the denominator of the ratio is readily measurable. But this velocity is not even conceptually the same as the ratio of income to currency in the observed economy. In the observed sector, cash is used in only a small fraction of the transactions, and the ratio of income to currency is nothing like total velocity of money in that sector. Cash is but a small part of total the money supply: in Australia in recent decades, currency has been only 6–9 per cent of M3 money (and even less of that other measure recently favoured by the makers of monetary policy, 'broad money').[14] Most transactions in the modern economy (at least in value) are not settled by cash, but instead by instructions to a banker in the form of a cheque, draft or electronic transfer. The dominance of the banks in the payment system is reflected in the relative amounts of currency and other forms of money in the community, in which there is many times more non-cash than cash.

[14] M3 is currency in circulation plus deposits with the banks by the non-bank private sector, including building societies and credit unions. Broad money is M3 plus the net deposit holdings of the non-bank institutions.

What is called velocity of currency in the observed economy in equation 10, is the total of all observed income in a ratio to the stock of 'legal' currency that is used in the observed economy. The ratio of observed NNI to total currency in the hands of the Australian public is in the range 21–26 over the period. So, if 15 per cent of the currency on issue is being siphoned off to fund underground economic activity, as indicated by the estimates, the implied velocity of 'legal' currency in the recorded sector will be even higher by that proportion, therefore in the range 24–30. This is the assumption of velocity implicit in the calculations of an underground economy that is 15 per cent of official GDP.

A valid calculation of velocity of currency in the observed economy, to match the velocity calculation in the underground economy, would be to take the value-added in all transactions that are actually settled in cash in a ratio to the stock of currency. Although this would obviously yield a much lower number than the previous ratio, the calculation is impractical because the numerator of the calculation is not measured. An alternative simple assumption of 'equal velocities' would equate the total flow of income in each sector to the total stock of *money* available that sector, not just currency money. In Australia, in recent decades the ratio of NNI to M3 money has been approximately 1.3–2.4, and more recently at the lower end of that range. (The effect of currency removed to the underground economy is negligible in this calculation.)

One way to justify this calculation is to imagine what the velocity of cash might be if there were only cash available for settling of transactions in an otherwise modern economy. A first approximation suggests that the velocity of cash in such a situation would have to be around the same level as the velocity of all money in the actual economy, that is, of the order of 2. It can be argued that number is somewhat too high or too low. Because cash is less efficient than cheques or electronic transfers for large or distant transactions, perhaps even more cash would be required to support the same level of activity in our imaginary economy than the total of all forms of money today. Thus, we might agree that 2 is perhaps too high a figure for the velocity of cash in a cash-only economy. On the other hand, some components of M3 are deposits that are not used in regular transactions, so on that account it is arguable that the figure is too low.

Velocity in the underground economy is similar in concept to our hypothetical cash-only economy, so a similar figure might be a good starting point for the value of velocity. Thus, we may conjecture a value for that velocity also, approximately 2. The effect of non-transaction motives for holding cash may be even more dramatic in the underground sector than in the legal economy. Persons engaged in illicit earnings will avoid purchasing real property or non-money financial assets because they fear detection through these dealings, so they may choose to hold more of their assets in the form of money. Because the only money available to them is currency, their non-transactions demand for cash will be higher than that of legitimate income earners, and relatively less of the cash will be available to finance their transactions. More cash will be needed to support the same level of activity in the underground economy, implying a still lower velocity in that sector.

Much of the other literature that uses currency demand modelling to estimate hidden incomes uses an 'equal velocity' assumption of some kind. However, in those cases the assumption is that velocity of currency in the underground sector is the same as the ratio of income to all forms of money in the observed economy. There are differences between these studies in what constitutes money and, in common with other monetary analysis, there is a general broadening of the definition with time. Thus, Tanzi (1983) and Schneider (1986) use the relatively narrow M1 measure in their studies of the USA and Denmark, respectively.[15] (Velocity is in the range 2–7 in the Tanzi study, but not reported by Schneider.) In a somewhat later study of Canada, Mirus *et al.* (1994) use the income velocity of an M2 measure of money for the same purpose.[16] More recently, Gadea and Serrano-Sanz (2002) prefer a money aggregate that is even broader than M3 in their study of Spain, and which has an average income velocity of 1.15. Both Klovland (1984) for Sweden, and Hill and Kabir (2000) for Canada, discuss the direct and indirect evidence for the velocity of currency in the underground economy of their respective countries, and both consider that a range of 2–7 will cover the possibilities. Most of the literature cited in these latter two studies argues for a broad definition of money in the observed economy to be used as the comparator to currency in the hidden sector, and, hence, it argues for values at the lower end of the velocity range, that is, values of approximately 2.

Against these considerations, the implied value of velocity of 24–30 in the Bajada estimates is seen to be many times too high. Because the estimate of the

[15] M1 is currency plus current deposits with the trading banks. This measure is not used much in Australia now because there is no clear distinction of a 'current' deposit or a 'trading' bank.

[16] M2 is M1 plus (trading) bank term deposits. This measurement has not been used in Australia for many years.

underground economy is in direct proportion to the assumed velocity, the estimate also is too high by the same factor. If a more sustainable value for velocity were to be used, the same calculations based on 'excess sensitivity' would yield an estimate of underground income, not 15 per cent, but instead approximately 1 or 2 per cent of observed GDP.

IX Conclusions

The method of 'excess sensitivity' used by Bajada (1999, 2001, 2002) indicates a large underground economy in Australia, with an estimate of hidden income of 15 per cent of GDP. But the method is unsound, because a simple change in the units of measurement will produce a completely different estimate. The effect of units is not trivial, because a rescaling of the tax variable from a percentage to a decimal fraction will cause the estimate of underground income to change from large and positive to large and negative (and, hence, be meaningless). The measure of 'excess sensitivity' can be immunised against changes in the units of measurement, but the respecification necessarily produces an estimate that is, at best, trivially small.

Some warning signs about the methodology can be seen in the empirical results. Despite its apparent complexity, the calculation is simple in essence: the estimate of the underground economy is almost entirely a mathematically transformed version of income tax payments relative to GDP. The calculation comes from static predictions using the fitted regression, with no reference to the dynamic properties of the error-correction model that is estimated. The published estimate has to be treated with a strong and arbitrary moving-average filter, without which the estimate is highly seasonal and erratic, even though the tax variable on which it is based is already deseasonalised in the data file.

Separately from the problems of measuring 'excess sensitivity', the value that is assumed for the income velocity of currency is many times too high to be plausible. The resulting estimate of underground economic activity in Australia is incredible on that account alone.

REFERENCES

Australian Bureau of Statistics. (2004), *The Underground Economy and Australia's GDP*, National Accounts Branch Discussion Paper, Australian Bureau of Statistics, Canberra.

Bajada, C. (1999), 'Estimates of the Underground Economy in Australia', *Economic Record* **75**, 369–84.

Bajada, C. (2001), *The Cash Economy and Tax Reform*, Research Studies Series No. 36, Australian Tax Research Foundation, Sydney.

Bajada, C. (2002), *Australia's Cash Economy: A Troubling Issue for Policymakers*, Ashgate, Aldershot.

Bajada, C. (2003), 'Business Cycle Properties of the Legitimate and Underground Economy in Australia', *Economic Record* **79**, 397–411.

Gadea, M.D. and Serrano-Sanz, J.M. (2002), 'The Hidden Economy in Spain – A Monetary Estimation 1964–1998', *Empirical Economics* **27**, 499–527.

Hill, R. and Kabir, M. (2000), 'Currency Demand and the Growth of the Underground Economy in Canada, 1991–1995', *Applied Economics* **32**, 183–92.

Klovland, J.T. (1984), 'Tax Evasion and the Demand for Currency in Norway and Sweden. Is There a Hidden Relationship?' *Scandinavian Journal of Economics* **86**, 423–39.

Mirus, R., Smith, R. and Karoleff, V. (1994), 'Canada's Underground Economy Revisited: Update and Critique', *Canadian Public Policy* **20**, 235–52.

OECD (Organisation for Economic Co-operation and Development). (2002), *Measuring the Non-Observed Economy: A Handbook*, OECD, Paris.

Schneider, F. (1986), 'Estimating the Size of the Danish Economy Using the Currency Demand Approach: An Attempt', *Scandinavian Journal of Economics* **88**, 643–68.

Schneider, F. and Enste, D.H. (2000), 'Shadow Economies: Size, Causes and Consequences', *Journal of Economic Literature* **38**, 77–114.

Tanzi, V. (1982), 'Underground Economy and Tax Evasion in the United States: Estimates and Implications', in Tanzi, V. (ed), *The Underground Economy in the United States and Abroad*, Lexington Books, Lexington.

Tanzi, V. (1983), 'The Underground Economy in the United States: Annual Estimates, 1930–80', *IMF Staff Papers* **30**, 283–305.

Verbeek, M. (2000), *A Guide to Modern Econometrics*, Wiley, Chichester.

Appendix I

Dr Christopher Bajada provided me with the file of data I use in this paper. I have not been able to check the data fully against the original ABS sources, because the national accounts have been revised extensively in recent years (including new methods of adjusting for price movements and a rebasing of the price indices), and the pages of AusStats have been rearranged. Most of the variables in the file are quarterly and seasonally unadjusted. One exception is the tax variable, which is not the ABS quarterly variable, but is instead constructed at quarterly frequency by taking one-quarter of the annual amount for the financial year. Another exception is the variable *GNE*, which appears to come from seasonally adjusted sources, although that variable is not used in the present paper. A copy of the data file has been lodged with the editors of the *Economic Record*. Some of the univariate results in Bajada (1999) can be replicated with the data in this file, but other calculations including the regression results in that paper require additional data.

Subsequent to submission of the final version of my paper, I received additional data from Christopher Bajada. These data have enabled me to replicate the basic regression results in Bajada (1999, table 3, p. 376).

The percentage tax and welfare variables used in the calculation of 'excess sensitivity' in the present paper are constructed from the variables in the file as follows:

$$T = 100 \times \left(\frac{Tax}{GDP} \right),$$

$$W = 100 \times \left(\frac{Welf}{GDP - Tax + Welf} \right),$$

where *Tax* is called *T* in the file, *Welf* is called *W*, and *GDP* is called *GDP(I)*.

A previous version of this paper was, of necessity, based on an extract of ABS data to match as nearly as possible the descriptions of data items given in Bajada (1999, Appendix). When I originally requested the data from the author, I was told they were lost. The editors of the *Economic Record* were unable to supply the original data, despite the paper that used it being published in a volume of the journal that bears inside its cover the statement

> It is the policy of the Economic Record to foster the replicability of empirical results by other researchers. Accordingly, the Economic Record normally publishes papers only where the data used in the analysis are clearly and precisely documented, are available to others to enable replication of results, and where details of the computations sufficient for replication are provided.

and which contains in its instructions to authors the requirement

> If the paper presents analysis of data, it should be accompanied by a computer disk with data files which enable a referee to replicate both the construction of the data and the results. A hard copy of the requisite documentation should be attached.

According to the current editor, the manuscript for the original paper would have entered the review process before the policy on submitting data was adopted. That is perhaps understandable in a period of transition, but misleading. Statements that in effect promote the quality of the journal's contents should apply to the current issue of the journal, not just to the unspecified future when the new policy becomes effective.