The monetary method to measure the shadow economy: The forgotten problem of the initial conditions

Hildegart Ahumada a, Facundo Alvaredo b,c,⁎, Alfredo Canavese a,c

a Universidad Torcuato Di Tella, Argentina
b Paris School of Economics, France
c CONICET, Argentina

ABSTRACT

We show that when the money demand function used to estimate the size of the shadow economy includes the lagged dependent variable, the need to assume a known initial condition reappears as it was the case in the early monetary methods.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The monetary method to measure the size of the shadow economy has become extremely popular, probably due to its presumed simplicity. The wide diversity of results obtained has generated skeptical views on its applicability.1 It has also received extrinsic critiques. On the theoretical side, its weak foundations have been stressed. On econometric grounds, researchers have criticized its quantitative accuracy due to time series properties, structural breaks and sensitivity to units of measurement.2 However, many papers that use the monetary method have been (and are still) published, certainly because the size of the shadow economy is a relevant figure for economic policy and because other methods are not necessarily better fit to deal with those critiques.3

The monetary approach is based on the assumption that cash is used to make transactions that agents want to keep hidden from official records. Transactions made using cash are difficult to trace, while those made with other assets, registered in financial institutions, can be easily inspected. If the amount of currency used to make hidden transactions can be estimated, then it could be multiplied by the income-velocity of money to get a measure of the size of the shadow economy.

The method was first presented by Gutmann (1977) and Feige (1979) and it evolved to use econometric tools in estimates made by Tanzi (1982), which are based on Cagan (1958).

Gutmann’s (1977) strategy requires four key assumptions: (a) high taxes and government regulations are the main causes of the shadow economy; (b) money is demanded for hoarding, transactions and precautionary motives; (c) transactions in the underground economy are made with cash; and (d) underground transactions are a relevant share of total transactions in the economy. The assumption of a known initial condition is one of the key assumptions in Gutmann’s (1977) strategy.

Gutmann (1977) also proposed an alternative model to estimate the size of the shadow economy by using a linear regression involving the money demand function. Gutmann’s model was later modified by Feige (1979). This model is the one used in this paper to estimate the size of the shadow economy.

1 See Tanzi (1999).
2 See Thomas (1999) and Breusch (2000a,b) among others.
3 Together with the monetary method, three other approaches coexist in the literature, based on: (1) registered income and observed expenditure; (2) direct auditing surveys; (3) declared and observed use of production factors. The different estimation methods target different concepts (income generated by underground activities and not reported to tax authorities, income not registered by national accounts, etc). However, as the very nature of the shadow economy makes its measurement a difficult task, the meaning of the estimates is generally unclear.
existence of a shadow sector; (b) only cash is used to make transactions in the shadow economy; (c) the ratio of currency to demand deposits, \( CD \), is only influenced by changes in taxes and regulations, and (d) there was some past moment in which no shadow economy existed. As the ratio \( CD \) of that period should have prevailed except for changes in taxes and regulations, each increase in \( CD \) is directly linked to the extra currency used in the shadow economy. The method assumes that the income-velocity of circulation, \( v \), is equal for the registered and the hidden economies; hence the size of the hidden sector is \( v \) times the extra currency.

Feige’s (1979) method uses the standard version of the quantity theory of money. If the ratio of the value of transactions to nominal income remains constant through time and is known for a period in which there were no hidden transactions, then total nominal income can be estimated for any period. The difference between estimated total nominal income and observed nominal income is the size of the shadow economy.

The work by Tanzi (1982), and all the papers based on his approach, claim that the income-velocity depends not only on variables that induce economic agents to make hidden transactions but also on income and the opportunity cost of holding cash. Demand for currency estimates are used to get the extra cash held by economic agents to finance hidden transactions. It is again assumed that the income-velocities of circulation for registered and hidden transactions are equal, so the size of the shadow economy is \( v \) times the extra cash.

In Ahumada et al. (2007) we showed that in almost all empirical applications the steps followed to “measure” the shadow economy are inconsistent with the assumption of equal velocities, unless the income elasticity of the currency demand is one. The technique is not only based on the fact that the velocity is lower in the shadow economy, but it also provides two distinct estimates, so that the assumption of equality is wrong and unnecessary.

Here we focus on another pitfall that has slipped into much of the applied work as well. The approaches suggested both by Gutmann (1977) and Feige (1979) made a “heroic” assumption: there is a period of time within the sample analyzed in which the black economy did not exist and all currency was demanded for legitimate transactions. Tanzi’s technique does not need this assumption, as he implicitly deals with the long-run: the variable which induces hidden transactions in the demand for currency is set to zero to get registered cash, and then extra cash is easily obtained. Nevertheless, as soon as short-run models of the demand for currency including the lagged dependent variable are introduced to “improve” the shadow economy estimation, the ad hoc assumption of a known initial condition comes back into play. Most published results are also flawed in this respect.

2. The monetary approach and the “long-run”

Here we review the monetary method within the aggregation framework shown in Ahumada et al. (2007). Eq. (1) displays the money demand function usually assumed in this approach

\[
C_{t1} = A(1 + \Theta_{t}^{\alpha}Y_{t}^{\gamma} \exp(-\gamma_{t})
\]

where \( C_{t1} \) denotes observed cash balances at time \( t \), \( \Theta_{t} \) is a variable which induces agents to make hidden transactions (as the ratio of taxes or government expenditure to GDP), \( Y_{t} \) is a scale variable (as registered GDP), \( \gamma_{t} \) measures the opportunity cost of holding cash (the interest rate or the rate of inflation); \( A, \alpha, \beta, \) and \( \gamma \) are positive parameters. Observed currency, \( C_{t1} \), is equal to total currency, \( C_{t0} \), which includes cash used for recorded transactions, \( C_{t0} \), plus cash used for hidden transactions, \( C_{t1} \),

\[
C_{t1} = C_{t0} = C_{t1} + C_{tH}
\]

Under the assumption that the demands for \( C_{t0} \) and \( C_{tH} \) follow the form suggested by Cagan (1958) with equal parameters, Eq. (2) can be explicitly aggregated as:

\[
C_{t0} = AY_{t}^{\beta} \exp(-\gamma_{t}) + AY_{t}^{\beta} \exp(-\gamma_{t}) = AY_{t}^{\beta} \exp(-\gamma_{t}) \left( 1 + \left( \frac{Y_{t}^{\beta}}{Y_{t}^{\beta}} \right) \right)
\]

Observe GDP, \( Y_{t0} \), is the registered GDP, \( Y_{t0} \), which does not include hidden GDP, \( Y_{tH} \), so

\[
Y_{t} = Y_{t0} + Y_{tH} = Y_{t0} + Y_{tH}
\]

Thus, the shadow economy size is obtained by setting \( \Theta_{t} \) equal to zero in Eq. (1) to get an estimate of the amount of cash demanded under no incentives to hide transactions, \( C_{t0} \).

\[
\hat{C}_{t1} = AY_{t}^{\beta} \exp(-\gamma_{t})
\]

As \( \hat{C}_{t1} \) is known from Eq. (5) and \( C_{t0} \), is observed currency, \( C_{t0} \), then \( \hat{C}_{tH} \) can be obtained by difference,

\[
\hat{C}_{tH} = \hat{C}_{t1} - \hat{C}_{t1}
\]

Then, the ratio between \( C_{t1} \) and \( C_{tH} \) is

\[
\frac{C_{t1}}{C_{tH}} = \frac{AY_{t}^{\beta} \exp(-\gamma_{t})}{AY_{t}^{\beta} \exp(-\gamma_{t})} = \left( \frac{Y_{t}^{\beta}}{Y_{t}^{\beta}} \right)
\]

Eq. (7) provides an expression for \( \hat{C}_{tH} \), given \( Y_{t0}, \hat{C}_{t1}, \hat{C}_{tH} \), and \( \beta \). It also shows how \( \beta \) matters for estimation. It should be recalled here that the starting point was Eq. (1), which can be seen as a long-run demand for currency.

3. Short-run vs. long-run estimates

During the last twenty years more accurate representations of money demand have been developed to take into account the time series properties of the data involved (cash holdings, income, interest rates, inflation), which are either integrated or highly persistent. From ‘partial adjustment’ to ‘equilibrium correction’ these dynamic models admit two types of estimates: one for the short-run and another for the long-run. Thus, long-run estimates (Eq. (1)) were replaced by short-run estimates as a central input to the method. When the short-run models that include the lagged dependent variable are used to obtain the size of the shadow economy, the problem of the initial conditions reappears.

Consider a usually estimated form of the currency demand, the simplest partial adjustment model:

\[
c_{t} = b_{1} + b_{2}c_{t-1} + b_{3}Y_{t} + b_{4}K + b_{5} \ln(1 + \Theta_{t})
\]

where lower case letters stand for logarithms. Eq. (8) can be seen as a model of partial adjustment to reach the equilibrium for desired real cash holdings

\[
c_{t} = \ln(A) + \beta y_{t} - \gamma_{t} + \alpha \ln\left(1 + \Theta_{t}\right)
\]

Eq. (9) is just the log of Eq. (1). The long-run solution is

\[
\frac{\hat{C}_{t1}}{\hat{C}_{tH}} = \frac{b_{1}}{1 - b_{2}} + \frac{b_{2}}{1 - b_{2}} y_{t} + \frac{b_{3}}{1 - b_{2}} k_{t} + \frac{b_{5}}{1 - b_{2}} \ln\left(1 + \Theta_{t}\right)
\]
and, comparing Eqs. (1) and (10).

\[
\begin{align*}
\frac{b_1}{1-b_2} &= \ln(A); & \frac{b_3}{1-b_2} &= \beta; & \frac{b_4}{1-b_2} &= -\gamma; & \frac{b_5}{1-b_2} &= \alpha \\
\end{align*}
\]

We have already shown in Section 2 that \(C_R\) can be directly obtained from Eq. (9) making \(\Theta_t = 0\); then Eqs. (6) and (7) give the size of the shadow economy. However, if Eq. (8) is used instead of Eq. (9) to obtain \(C_R\), a new key problem arises: \(C_{Rt-1}\) (or a previous value, or the initial condition) is required. Carelessly, in applied work, \(C_{Rt-1}\) has been used instead of \(C_{Rt-1}\). Then, the size of the shadow economy is wrongly estimated.

To further analyze this issue it is convenient to show the kind of aggregation assumed in the partial adjustment model. Eq. (8) is obtained by taking logs in

\[
C_{Rt} = C_{Rt-1}^{\lambda} C_{t}^{(1-\lambda)} = (C_R + C_{t-1})^{\lambda} (C_R + C_{t-1})^{(1-\lambda)}
\]

where an asterisk denotes the long-run equilibrium.\(^8\)

When \(\Theta_t = 0\), only the part of \(C_{Rt}\) due to registered transactions can be obtained, but not \(C_{Rt}\). \(C_{Rt-1}\) (unknown) is required to get \(C_{Rt}\). When \(C_{Rt-1}\) is used instead, \(C_{Rt}\) is overestimated because \(\lambda\) is positive. If \(C_{Rt-1}\) is substituted by \(C_{Rt-2}\), and \(C_{Rt-2}\) by its previous value and so on, the required knowledge about registered currency would be moved eventually to an initial value \(C_{R0}\) as can be seen by solving the pertinent first difference equation for the log of currency demand.

4. Conclusion

If short-run representations of the money demand that include the lagged dependent variable are used to estimate the size of the shadow economy, the need to assume a known initial condition reappears (as it was the case in the early literature). Consequently, the only way to avoid ad hoc assumptions about previous values of registered currency is to restrict the measures of the shadow economy size to those based on the long-run estimates of the money demand.

References

Breusch, T., 2005b. ‘Fragility of Tanzi’s method of estimating the underground economy’, The Australian National University, mimeo.