Chapter 12

MACRO-MODEL METHODS

12. MACRO-MODEL METHODS

12.1. Introduction

12.1. *Macro-model methods* is the term used in this Handbook for methods that produce an estimate of the entire NOE, or one of its component parts such as the underground economy, by means of a single model. Such methods are discussed in this chapter not because they are considered useful in obtaining exhaustive estimates of GDP or in estimating underground production, but because they tend to produce spectacularly high measures, which attract much attention from politicians and newspapers. As Gylliane Gervais (Statistics Canada, 1994) wrote:

> "The size and growth of the "underground economy' have kindled a lot of interest in Canada in the past few years…Hardly a week goes by without the media quoting someone claiming that underground transactions amount to 10%, 15% or even 20% of GDP, or that the deficit could be eliminated if taxes were collected on these transactions. If the figures often quoted are even approximately accurate, then the level and possibly growth of Canada's GDP are considerably understated to the extent that the information given (to) policy makers about current economic conditions is misleading. Our statistical system, at least in so far as economic statistics are concerned, would be sorely inadequate if it failed to detect "hidden' transactions of such magnitude."

12.2. Without a doubt, the media reports to which this paragraph refers were based on the use of macro-model methods. For example, Schneider and Enste (2000) quote estimates of the average level of the "shadow economy" (apparently meaning non-measured economy) in Canada over the period 1990-93 between 10.0% and 13.5%, whereas the Statistics Canada report just referenced concluded that the upper limit to the "underground economy" (defined as market production of legal goods and services that escapes measurement in the official GDP estimates) in 1992 was 2.7%.

12.3. It is often, but entirely wrongly, conjectured that the difference between these macro-model results and the official estimates of GDP is non-measured production. However, as illustrated in this chapter, there are serious problems with macro-model methods that cast doubt on their utility for any purpose in which precision is important. In particular, they are completely unsuitable for use in compiling the national accounts. In brief the problems are:

- the activities that the models aim to measure are not precisely defined; it is often unclear whether the models are estimating non-observed or non-measured production, or whether they include informal sector or illegal activities as well as underground activities;
- the assumptions underlying the models overly simplistic;
- the results the models produce are not stable in the sense that changes in assumptions for the same model can produce quite different values;
- there are many models and they give different results;
- the methods provide only a global estimate for the economy as a whole, whereas users frequently want GDP broken down by industry or expenditure category; and

- the results can not be readily combined with other measurements, in particular those obtained by the detailed, data based methods that are used in compiling the national accounts, as described in earlier chapters of this Handbook.

12.4. To illustrate these points, three types of macro-model methods are described in the following sections:

- M*onetary methods*, which assume that the non-measured production can be modelled in terms of stocks or flows of money.
- G*lobal indicator methods*, in which non-measured production is modelled in terms of a single variable (usually a physical indicator) with which it is believed to be highly correlated, electricity consumption being the most commonly used.
- L*atent variable methods*, in which modelling is in terms of two groups of variables, one group that is assumed to determine the size and growth of non-measured production and a second group that provides the "trace" (*i.e.*, evidence) of the missing activities.

12.5. In presenting their methods, the authors refer variously to the "underground", "hidden", or "shadow" economy without always clarifying whether this includes all non-observed activities, or just those that are legal, or those that are non-measured in the official GDP estimates, or some other subset. The following descriptions should be read with this in mind.

12.2. Monetary Methods

12.2.1. *Introduction*

12.6. Three monetary methods are described in the following paragraphs, namely the *transaction method*, the *cash/deposit ratio* method, and the *cash demand* method. The fundamental idea behind these methods is to build relationships between monetary developments and official GDP estimates using regression techniques with a few restrictive hypotheses concerning the cash character of underground transactions, tax burdens, the relationship of holdings to deposits, etc., and to assume that all monetary developments not explained by the particular model are due to undercoverage of the official GDP.

12.2.2. *Transaction Method*

12.7. The transaction method described by Feige (1979) is based on the following reasoning. The starting point is the equation of Fisher:

$$M * V = P * T$$

12.8. In words, the total stock of money (M) multiplied by the velocity of circulation (V) equals the total number of transactions paid by that money (T) multiplied by the price of these transactions (P). Further it is assumed that there is a constant relationship (denoted by k) between the money flows related to these transactions and total value added, that is:

$$P * T = k * Ytotal,$$

where, by definition, total value added (Ytotal) is the sum of the official valued added (Yofficial) and underground value added (Yunder). Therefore

$$M * V = k * (Yofficial + Yunder),$$

thus $M_t * V_t = k * (Yofficial_t + Yunder_t)$, over some range of years $t = 0,1,\ldots T$.

12.9. The stock of money (currency plus demand deposits) is readily measurable, money velocity can be estimated and the official estimates of value added are known. Thus, if the size of the underground economy as a ratio of the official economy is assumed known for a benchmark year, then the underground component can be calculated for all subsequent years.

12.10. Feige applied the method to the United States. Following Laurant (1979), he estimated the velocity of cash as the quotient of the number of transactions a bank note survives before being worn out and the average lifetime of banknotes. The first part of this estimate was based on bank

note wearing tests. He then assumed that the underground economy was zero in 1939. Based on these assumptions he calculated the size of the underground economy as 27% in 1979.

12.11. There are several problems with respect to Feige's method. The assumption of a constant ratio of transactions to official GDP seems heroic. As Cramer (1980) pointed out it is highly likely that some monetary transactions that have nothing to do with income generation are included in the calculations, for example the results of the introduction of a stricter cash management in large companies and of repurchase agreements and of euro-dollar deposits, which very frequently change ownership in the United States. Evidently part of the money notionally in circulation, in particular bills of large denomination, is not actually in circulation but kept as a store of wealth. Furthermore, the amount of money held as cash depends upon interest and inflation rates and people's perception of the likelihood of being robbed. Increased facilities for and use of cheques and credit cards can also be expected to have had an impact.

12.12. Cramer also criticised Feige's estimate of the income velocity of cash money. Cramer preferred an alternative estimate based on the number of cash withdrawals at banks and the average number of uses of a banknote between withdrawal and deposit.

12.13. Blades (1982) pointed out that the US dollar is an international currency unit and that US dollars circulate widely throughout the world, either as official legal tender (in the Virgin Islands, Liberia, Puerto Rico and Samoa for example) or as a widely accepted alternative to national currencies that lack credibility (in South East Asia and Central America for example). Because the dollar plays an international role, there is little point in relating dollars in circulation (throughout the world) to domestic activity in the United States.

12.14. Applied to the Netherlands, the transaction method yields implausible results. To address this problem, Boeschoten and Fase (1984) refined the transaction method, creating a new base method and several other variants. However, the variants all give significantly different results and there is no way to deduce which might be closest to the truth.

12.2.3. *Cash/Deposit* Ratio Method

12.15. The cash/deposit ratio method is based on information about the ratio between currency (cash) and transferable (giro) money, *i.e.*, the *cash/deposit ratio*. According to Gutmann (1977) the cash/deposit ratio is only affected by the changes in taxation and other government regulations that change the way people make payments, and the main reason that payment behaviour changes is because people want to hide certain activities to avoid taxation and restrictions. Gutmann refers to these hidden activities as the underground economy and uses the cash/deposit ratio to estimate its size.

12.16. Gutmann illustrates the method by an example from the United States. After a short rise just after the Second World War, the cash/deposit ratio for the United States remained almost constant until 1961, when it began to increase again. According to Gutmann, this conflicted with the developments that occurred before the war. Given that he expected both the pre-war and post-war developments to be the same, he concluded that the increase of the cash/deposit ratio from 1961 was the result of changed behaviour. He conjectured that an increasing proportion of transactions was being paid in cash, which pointed to the creation or increase of the underground economy. Based on the changes in the cash/deposit ratio Gutmann computed the underestimate of GNP because of the underground economy as 10% in 1976. The assumptions underlying this estimate are as follows:

- The cash/deposit ratio of the official (observed) economy has not changed since 1937-1961.
- Transferable money was only used in the official economy.
- The "surplus" of money held as cash is only used in the underground economy. This surplus is calculated as the difference between the actual amount of cash money in circulation and the amount that one would expect to be in circulation according to the cash/deposit ratio of the official economy over the period 1937-1961.

- One dollar money held in cash in the underground economy generated as much value added as one dollar of M1 (cash + transferable) money in the official economy.
- The underground economy over the period 1937-1961 was negligible.

12.17. Many people have developed variants of the cash/deposit ratio method. Feige (1980), for example, changed the assumption of an almost non-existent underground economy in the benchmark years 1937-1941. He assumed an underground economy of 5% of GNP in 1964. Based on this new assumption he arrived at an estimate of 14% of GNP in 1979. Furthermore he challenged the assumption that one dollar in the underground economy was as productive (in terms of generated GNP) as one dollar in the official economy. Because of the higher share of services in the underground economy, he conjectured that one dollar in the underground economy was at least 10% more productive. Furthermore he wondered why all underground transactions should be in cash and assumed that about one third of all underground transactions were paid through bank transfers. Based on these assumptions he estimated the size of the United States underground economy in 1979 at 28% of GNP.

12.18. Other people have criticised the underlying assumption that changes in the cash/deposit ratio are due only to the underground economy. For example, Garcia (1978) put forward different possible reasons for changes in the cash/deposit ratio. In his view, the growth of the underground economy was not the only reason, nor even the most important reason. More important were the decrease of money in the form of checking accounts and the increase of various types of savings accounts. If corrected for this shift, the cash/deposit ratio in the United States turns out to have been rather constant.

12.19. Laurent (1979) and Cramer (1980) also regarded the cash/deposit ratio as an inappropriate means of measuring the underground economy. In their view, the velocities of circulation of the different types of money develop differently, thus a better measure is the total amount of cash and bank transfer payments. This was the reason they preferred the transaction method previously outlined.

12.2.4. *Cash Demand* Method

12.20. In contrast to Gutmann, Tanzi (1980, 1982) assumed that the demand for cash money was not only affected by taxation and government regulation, but also by other factors. However, he agreed with Gutmann that changes in the total amount of cash money due to changes in taxation and government regulation go totally into the underground economy. To isolate the influence of taxation and regulation Tanzi postulated that the demand for cash as a proportion of total money, C/M2 (where M2 is cash money + transferable money + fixed period deposits), is a function of taxes, the share of wages and salaries in total personal income, the interest on fixed term deposits, and per capita real income. Based on the results of regression analyses, Tanzi arrived at two alternative estimates of the *notional demand* for cash money (defined as the demand for cash money on the assumption that there is no underground economy). They were the notional demand if taxes were unchanged since 1929 (the year when Tanzi assumed there was no underground economy), and the notional demand if there were no taxes.

12.21. In each case, the difference between the actual demand and the notional demand was considered to be the total amount of cash money in the underground economy. Assuming that the velocities of cash money in the formal and the underground economy are equal, the estimated size of the underground economy in the United States in 1976 was 3.4-5.1% of GNP according to the first variant and 8.1-11.7% according to the second variant. These values are both quite different from Gutmann's estimates previously described.

12.22. Barens (1982) demonstrated that both cash/deposit and cash-demand methods can yield counter-intuitive results. He applied both methods to the Netherlands and the results showed a decreasing underground economy since the end of the seventies, which is in contradiction to the generally accepted notion that the underground economy was increasing.

12.2.5. *Summary*

12.23. Monetary macro-models are unsuitable for estimating the underground economy primarily because they are based on modelling assumptions that cannot be justified. The crucial assumption

underlying the cash/deposit ratio and the cash demand methods is that a change in the size of the underground economy is caused by changes in taxation and government regulations and that this becomes visible through changes in the demand for cash because underground transactions are mainly paid in cash. This assumption cannot be tested and may not be true. In contrast, the transaction method is not based on any assumed relation between taxation and underground activity. However, it assumes a constant relationship between monetary transactions and GDP, which again cannot be justified.

12.24. The problems with these models are evidenced by the sensitivity of the results to the benchmark year assumptions, the wide range of results that different methods give under the same circumstances, and the implausible results that are obtained in some specific cases.

12.3. Global Indicator Methods

12.25. The most prominent example of the global indicator approach is the electricity consumption method proposed by Kaufmann and Kaliberda (1996). This method uses electricity consumption as the single physical indicator of overall economic activity. It assumes a precise and stable relationship between electricity consumption and output. It is clear, however, that at least for agriculture the relationship will not be stable since output is largely determined by the weather. Moreover, in many developing and transition countries, electricity is not a major source of energy in industrial production.

12.26. The problem with the method is illustrated by an example drawn from the Russian Federation. In the years immediately following the beginning of intense economic reform in 1992, official statisticians experienced difficulties in producing sufficiently timely assessments of GDP, which was known to be in decline. Electricity consumption was one of the alternative methods tried. It produced more optimistic estimates than the official GDP and provoked considerable discussion of the NOE. Subsequent analysis indicated that these estimates could not be trusted for the following reasons.

- The relationship between electricity consumption and industrial production is not a simple ratio, even in those industries that are highly dependent upon it, because a significant part (up to a third) of consumption is a fixed cost unrelated to the volume of production. For example, factories need heating and lighting. This sort of consumption tends to change in steps rather than move smoothly with production.
- Artificial price levels further reduce the relationship between industrial production and electricity consumption. At this time in the Russian Federation, electricity prices were artificially low, thus manufacturers did not need to worry greatly about electricity consumption, nor to make efforts to reduce it when production fell. By the same token, they did not need to increase consumption much if production picked up again, either officially or through undeclared use of facilities.
- Measurements of consumption are actually derived from estimates of electricity production at the power stations. The difference between supply and consumption – transmission losses – is not large but can vary with the seasons.

12.27. Based on these arguments, the electricity consumption method was not used in the fundamental revision of the GDP undertaken by Goskomstat of Russia (1995) jointly with the World Bank. This example is typical of the problems encountered in considering such a simplistic approach to NOE measurement.

12.4. Latent Variable Method

12.28. The macro-modelling methods described in the previous paragraphs assume that the underground economy can be modelled in terms of a small number of specific variables. They ignore other background information and circumstances that can lead to the existence of underground production. This is not the case for the *latent variable method* of Frey and Weck (1983), which draws on a wide range of explanatory variables. The size of the underground economy is estimated on the basis of developments in the variables that, on the one hand, affect the size and the growth of underground production, and, on the other hand, are the traces of underground activities in the economy. The

method uses a technique (known as LISREL) that enables a cross-sectional analysis of the relationship between a non-observed dependent variable and one or more observed explanatory variables. As the non-observed variable is not known it is replaced by a set of indicators. The data are drawn from a range of different countries, or time periods within a country. The results of the analysis are estimates of the relative sizes of the non-observed variable in each of the countries, or time periods. To estimate the actual sizes, benchmark estimates for two countries, or time periods are required.

12.29. According to Frey and Weck, the size of the underground economy can be explained in terms of the actual tax burden, the perception of the tax burden, the unemployment rate, the regulation burden (for example, the number of laws), the attitude towards paying taxes (*tax morality*) and the per capita available income. Indicators of the size of underground production, in other words the traces left in the economy, are the labour force participation rate of the male population, the number of weekly working hours and the growth of the GNP. To build their model Frey and Weck used data from a large number of countries for a particular year. They calculated the relative sizes of the underground economy in these countries and then generated estimates of the actual sizes, using monetary method estimates for Norway and Sweden as benchmarks. Giles (1999) describes another application, in this case for New Zealand over a range of years.

12.30. The method can be seriously questioned. First, one can ask about the variables chosen. For example, why is the number of weekly working hours an indicator (trace) of the underground economy? Is it a reason for the existence of the underground economy rather than a result? Second, an important variable in the model is the so-called tax morality, but this is difficult to quantify in an objective fashion. Third, the reliability of the results is never better than the reliability of the two benchmark estimates. Fourth, the results are very unstable. Helberger and Knepel (1988) show that even a small change in the countries used in the Frey and Weck example leads to quite different results. The exclusion of Finland – a country which accounts for less than 1% of the population and of the GDP of the group of countries studied by Frey and Weck – leads to insignificant estimates of almost all coefficients in the model. Helberger and Knepel conclude that the shortcomings and ambiguities in the data severely limit the utility of the model for measurement of the underground economy.

12.5. Concluding Remarks

12.31. The need for models in compiling national accounts is not in dispute. Models often underpin indirect methods used in compilation, providing estimates when basic data are not available. This is usually the case, for example, in making estimates of illegal production. The point is, however, that the preferred basis for statistical estimates is always empirical data. Only where data are not available should modelling be used to fill the gaps; and the models should be at the most detailed level possible so that they have the best possible chance of accurately reflecting the phenomena being modelled. The goal should be to estimate each specific, non-observed data element using closely related data for the same, or a nearby, accounting period. National accounts compilation should follow a careful, case by case approach, considering all available data sources and procedures. In this context, an electricity consumption model, for example, may have its place, not in modelling the growth of the entire economy, but at a detailed level, such as estimating the growth of a particular industry for which survey data are not available and for which electricity is a primary input and reflects the movement in output.

12.32. In most countries, national accountants have available a rich supply of data from a variety of sources. In combination, these data are capable of producing much more accurate estimates of GDP and its components than macro-models can ever do. It is, however, incumbent on national accountants to inform users of the extent of the *non-observed economy – i.e.*, how much economic activity escapes direct measurement – and the extent of the *non-measured economy – i.e.*, how much of the non-observed economy may still be missing from GDP after making the various adjustments of the kind described in this Handbook. Lack of transparency in describing the procedures used to compile the national accounts is the main reason why outsiders resort to macro-models and produce estimates that undermine the credibility of the national accounts.