Estimating the Underground Economy using MIMIC Models

Trevor Breusch[1]

Abstract

MIMIC models are being used to estimate the size of the underground economy or the tax gap in various countries. In this paper, I examine critically both the method in general, and three applications of the method by Giles and Tedds (2002), Bajada and Schneider (2005) and Dell'Anno and Schneider (2003). Connections are shown with familiar econometric models of linear regression and simultaneous equations. I also investigate the auxiliary procedures used in this literature, including differencing as a treatment for unit roots and the calibration of results using other data. The three applications demonstrate how the method is subjective and pliable in practice. I conclude that the MIMIC method is unfit for the purpose.

Keywords: underground economy, MIMIC, structural modelling, LISREL® software.

JEL Codes: C22, C51, E26, H26.

INTRODUCTION

By definition, the underground economy cannot be directly observed so its magnitudes must be estimated. Many different methods are employed for this purpose. Tax audits are informative, but are usually targeted toward suspected offenders, and hence are biased estimators of aggregate behaviour. Regular surveys of household expenditures and incomes conducted by national statistical agencies can be examined for discrepancies that might indicate unreported incomes. Special surveys are sometimes conducted, with direct questions about below-the-counter incomes or cash payments, although non-response bias is always a concern. At a more aggregate level, inferences can be made from inconsistencies between expenditure, income and product data collected from various sources for national accounting purposes. The most popular methods in the academic literature are based on macroeconomic models of either the demand for currency holdings (perhaps in comparison with bank account balances) or the consumption of some standard commodity such as electricity.

Interest is burgeoning in a more complex approach known as the "structural equation" or "multiple indicator multiple cause" (MIMIC) model. This method has its origins in the factor analysis literature of psychometrics, while its exposure in economics has been through the latent variable models of Zellner (1970) and Goldberger (1972). In the first application of MIMIC to estimating the underground economy, Frey and Weck-Hannemann (1984) examine a pooled dataset from 17 OECD countries. The idea is extended by Aigner, Schneider and Ghosh (1988), who allow some lagged adjustment in a dynamic MIMIC (or DYMIMIC) model and apply the method to the US. Giles (1999) further modifies the approach to incorporate developments in time-series methods, especially unit roots and cointegration analysis, and provides estimates of New Zealand's hidden economy. The state of the art of dynamic MIMIC modelling is given by Giles and Tedds (2002), who describe the approach in detail and apply it to Canada. Authors

---

[1] Crawford School of Public Policy, The Australian National University

taking up the method in the wake of Giles and Tedds include Bajada and Schneider (2005), who study Australia and other Pacific nations, and Dell'Anno and Schneider (2003), who estimate the underground economy in Italy and report results for other OECD countries.

The MIMIC approach is attractive in this context. The idea is to represent the output (or income) of the underground economy as a latent variable or index, which has causes and effects that are observable but which cannot itself be directly measured. Thus, there are two kinds of observed variable in the model, "causal" variables and "indicator" variables, which are connected by a single unobserved index. Values of the index over time are inferred from data on causes and indicators by estimating the statistical model and predicting the index. The fitted index is then interpreted as a time-series estimate of the magnitude of the underground economy. Usually, the measure is hidden output or income as a percentage of recorded GDP, although some researchers are concerned with the "tax gap" between actual and potential revenue when all taxable income is reported.

Bold claims are made by proponents of these methods for their ability to measure hidden economic activity. The estimates in the literature are often presented to three or four digits of precision and without any interval of uncertainty. The estimates are always large enough to cause grave concern and attract media headlines, and often the underground economy is shown to be growing strongly. These results have serious implications for economic and social policy in the areas of tax administration, national income accounting, stabilization policy, and social fairness and cohesion.

This use of MIMIC modelling has its critics. Helberger and Knepel (1988) show that the pioneering results of Frey and Weck-Hannemann are unstable in the face of minor changes in either the data period or the group of countries studied. They also argue that the lists of causal and indicator variables are unconvincing for the purpose. Smith (2002) and Hill (2002) criticise Giles and Tedds' (2002) modelling, especially in terms of the absence of economic theory to guide the specification and the complexity of the estimation strategy. Echoing Helberger and Knepel's (1988) critique, they also question the relevance of the causal and indicator variables that are employed. Giles and Tedds' specification and results are examined more closely by Breusch (2005a), who shows that the time path of their estimate for Canada has little to do with any underground activity, but mostly reflects price inflation and real growth in the observed economy. Moreover, the level of their estimate is a numerical accident with no connection to any evidence in the data.

My objective in the present paper is to look more broadly at MIMIC modelling as it is employed in this literature. A three-way distinction can be made between the method itself, the various ancillary treatments, such as data transformations and the post-model calibration that is called "benchmarking", and the modelling decisions made when applying the method to a particular dataset. My starting point is to connect the method with the standard econometric models of linear regression and simultaneous equations. Much of the novelty in the MIMIC approach will be seen to reside in the labelling and interpretation of the calculations. The novel terminology and unfamiliar perspective are fostered by the adoption of specialist software packages such as LISREL® and Amos™. In most cases, exactly the same calculations can be described in terms that will be more familiar to the practising economist.

As examples of the method, I shall examine the three recent works mentioned above: Giles and Tedds (2002), Bajada and Schneider (2005) and Dell'Anno and Schneider (2003). In each case, I can replicate the MIMIC estimation results and the major inferences using both LISREL® and standard econometric software. There is considerable divergence in practice between the three applications, particularly in their interpretations of the latent variable and in their approaches to calibration and other adjustments. In every case, I discover undocumented transformations of the data, and speculate that the authors are unaware of making such transformations. As a result of these ancillary treatments, it is not always clear to the reader how, and by how much, the results of the MIMIC model are stretched and squeezed to fit some outside evidence.

I find instances where an inference about underground activity is sensitive to the units of measurement, so different substantive answers can be obtained simply by measuring the variables in different units. Sometimes this problem arises because of the form of calibration employed. In other cases, the dependence on units can be attributed to undocumented transformations of the data. Such sensitivity is an undesirable property in any measuring instrument, because the resulting measurement can be varied by changing a setting that is perceived to be irrelevant. The upshot is a method that lacks objectivity because it is open to manipulation and misrepresentation.

I examine critically the strategy of data differencing adopted in this literature to deal with unit roots and cointegration. The purpose of differencing is not always clear, but I show that the treatment is not an effective solution to any problem that matters, and may, in fact, cause serious problems. Independent of the issues of dynamic specification, the very idea of the underground economy as a latent variable is questionable. I provide evidence to show that the MIMIC model has precise statistical implications that are absent from this area of application.

In addition to the general principles examined in the main part of this paper, I have discovered many errors and anomalies while replicating the three studies. These additional findings are not essential to understanding the MIMIC method in general or its potential for estimating the underground economy, so they are gathered into an Appendix. However, this material does demonstrate some of the pitfalls that await users of the method, and contains important advice for readers who seek to interpret or employ the substantive results of the three studies.

MIMIC AND ECONOMETRIC MODELS

The MIMIC model is described by Giles and Tedds (2002, Ch.6) as a relationship between a vector y (p x 1) of indicator variables and another vector x (q x 1) of causal variables. These are connected by an unobserved latent variable η (scalar) as follows:

$$y_t = \lambda \eta_t + \varepsilon_t \tag{1}$$

$$\eta_t = \gamma' x_t + \xi_t, \tag{2}$$

Where γ (q x 1) and λ (p x 1) are unknown parameter vectors. The error terms ε_t (p x 1) and ξ_t (scalar) are assumed to have zero means, variances $\Theta = diag(\theta_1, \ldots, \theta_p)$ and ψ, and to be uncorrelated with each other. A model consisting of (1) and (2) cannot determine the scale of all of the parameters, so a normalization condition is required. There are many possibilities, but

Giles and Tedds adopt the convention of setting the first element of λ to be unity, as $\lambda_1 = 1$. The data are a time series of observations $t = 1,\ldots,N$. Estimation is typically by maximum likelihood, on the additional assumption that the error terms ε_t and ξ_t are jointly normally distributed and independent over time.

In the MIMIC model, x is weakly exogenous in the sense that all of these distributional statements are conditional on x. Thus, the model implies particular structures for the conditional mean and variance of the observed variables:

$$E(y_t|x_t) = E\left[\lambda(\gamma'x_t + \xi_t) + \varepsilon_t|x_t\right] = \lambda\gamma'x_t, \tag{3}$$

$$\text{var}(y_t|x_t) = \text{var}\left[\lambda(\gamma'x_t + \xi_t) + \varepsilon_t|x_t\right] = \text{var}\left[\lambda\xi_t + \varepsilon_t|x_t\right] = \lambda\lambda'\psi + \Theta \tag{4}$$

These results can be written as a reduced form regression equation

$$y_t = \Pi x_t + v_t \tag{5}$$

where $\Pi = \lambda\gamma'$ and $v_t \sim (0,\Omega)$, and where $\Omega = \lambda\lambda'\psi + \Theta$. In general, the structure of the MIMIC model will imply restrictions on the reduced form parameters Π and Ω.

I wish to consider in more detail the case of two indicator variables, $p = 2$, since that is the nature of all three applications to be examined. In detail, then:

$$\lambda = \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}, \text{ so } \Pi = \lambda\gamma' = \begin{bmatrix} \gamma' \\ \lambda_2\gamma' \end{bmatrix}, \Theta = \begin{bmatrix} \theta_1 & 0 \\ 0 & \theta_2 \end{bmatrix} \tag{6}$$

It can be seen that the reduced form has $2q + 3$ parameters ($2q$ elements in Π and 3 more in Ω). However, the underlying model has $q + 4$ parameters (γ, λ_2, ψ, θ_1, θ_2). When $q>1$, as is typical, the reduced form will be restricted by the model. Writing out the restricted reduced form equations in full gives:

$$y_{1t} = \gamma'x_t + v_{1t} \tag{7}$$

$$y_{2t} = \lambda_2\gamma'x_t + v_{2t} \tag{8}$$

where

$$\text{var}\begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} = \begin{bmatrix} \psi + \theta_1 & \lambda_2\psi \\ \lambda_2\psi & \lambda_2^2\psi + \theta_2 \end{bmatrix} \tag{9}$$

Here, the coefficient vector in the second equation (8) is in constant proportion to the coefficient vector in the first equation (7). There is no additional restriction on the variances in this case

since, given λ_2, there are three distinct elements to the variance matrix and three parameters (ψ, θ_1, θ_2).[2]

It is useful to write out the structure in the standard econometric form of a simultaneous equation model for this leading case of $p = 2$ indicators. Multiply (7) by λ_2 and subtract the result from (8),

$$y_{2t} - \lambda_2 y_{1t} = v_{2t} - \lambda_2 v_{1t} = u_t \quad \text{(say)} \tag{10}$$

which gives the model as

$$y_{2t} = \lambda_2 y_{1t} + u_t \tag{11}$$

$$y_{1t} = \gamma' x_t + v_{1t} \tag{12}$$

where (12) is just a repeat of (7). This is formally identical to a two-equation linear simultaneous model, with two endogenous variables and q exogenous variables. In fact, maximum likelihood here defines the standard econometric procedure of limited information maximum likelihood (LIML) on equation (11), because the second equation is already in reduced form, and the covariance matrix between u_t and v_{1t} is unrestricted. The last point is seen here

$$\text{var}\begin{bmatrix} u_t \\ v_{1t} \end{bmatrix} = \text{var}\begin{bmatrix} v_{2t} - \lambda_2 v_{1t} \\ v_{1t} \end{bmatrix} = \begin{bmatrix} \lambda_2^2 \theta_1 + \theta_2 & -\lambda_2 \theta_1 \\ -\lambda_2 \theta_1 & \psi + \theta_1 \end{bmatrix} \tag{13}$$

which has three free elements and determines three parameters, given that λ_2 is determined as a coefficient.[3] Thus, the indicator and causal variables of the MIMIC model match exactly the endogenous and exogenous variables of econometrics. The measurement equations in the MIMIC model define the structural relationship by which the endogenous variables are jointly determined in the model.

While the MIMIC model can be interpreted and estimated as a standard econometric model of linear simultaneous equations, it retains its other interpretation in terms of the latent variable. The variance parameters in the MIMIC model can be solved from the econometric model as follows:

$$\theta_1 = -\text{cov}(u_t, v_{1t})/\lambda_2 \tag{14}$$

$$\theta_2 = \text{var}(u_t) - \lambda_2^2 \theta_1 \tag{15}$$

$$\psi = \text{var}(v_{1t}) - \theta_1 \tag{16}$$

---

[2] When there are more than two indicator variables in the model, so that $p > 2$, there are restrictions on the covariance matrix as well as among the coefficient vectors of the restricted reduced form.

[3] LIML recognises the structure of one equation of a system, while treating the other equations in their reduced form and ignoring any covariance matrix restrictions. It therefore uses the same information about the structure of the model as two-stage least squares, to which it is asymptotically equivalent. With more than two indicator variables in the MIMIC model, maximum likelihood estimation is not simply LIML, because the restrictions on the covariance matrix would be ignored in LIML.

Since these are variances, all three must be positive for the estimate to be admissible as a MIMIC model, but LIML estimation will not impose non-negativity on the solution, and it is not difficult to construct examples in which any one of the implied variances is negative.[4] This is no different from the LISREL® and Amos™ packages, which by default do not restrict the variance estimates to being positive, although in some cases a warning message is issued when the estimated variance matrix is not positive definite.

One virtue of our interpretation of the MIMIC model as simultaneous equations is that it can be estimated without the specialist LISREL® or Amos™ software. Economists, who may be unfamiliar with such software and its conventions, can then see clearly what computations are being performed on their data. Often the simplest and most insightful way to apply LIML estimation is to recognise its equivalence to iterated generalized least squares (or Aitken) estimation for seemingly unrelated regression (see Pagan, 1979). This GLS procedure is available in many packages, such as the "sureg" command in Stata™. Of course, iterated GLS only yields directly the estimates of λ_2 and γ in (11) and (12), and perhaps the variance matrix in (13). Estimates of the other parameters in the MIMIC model can be recovered easily: the variances θ_1, θ_2 and ψ come from substituting the GLS variance and coefficient estimates into expressions (14)-(16).

The main use of the MIMIC model in this literature is to extract the latent variable η_t, which, in some sense, is interpreted as measuring the size of the underground economy. Since

$$E(\eta_t | x_t) = E(y_{1t} | x_t) = \gamma' x_t \tag{17}$$

the estimate of the latent variable is the predicted value of the first indicator variable (the one which is normalized to have unit coefficient in λ). Note that the prediction is made from the *restricted* reduced form, which will be estimated by LIML or GLS. The MIMIC model defines a proportionality relationship between the vectors of coefficients in the two reduced-form equations, so the prediction of the other indicator variable is just a rescaled version of the prediction of the indicator variable on which the normalization is made, where the factor of proportionality is the estimate of λ_2. By the invariance of maximum likelihood estimation, it makes no difference in principle which indicator variable is chosen for normalization, since the same estimates are defined, apart from the obvious change in scale. However, there are two important consequences of the normalization that should be considered: one is practical and the other may be important for interpreting the results.

In practice it is likely that one unrestricted reduced-form equation will fit the data much better than the other when estimated by OLS; in the language of instrumental variables, the exogenous variables may be much better instruments for one of the endogenous variables than the other. In that case, the restricted LIML estimates of the reduced-form coefficients will more closely resemble the unrestricted OLS estimates of the equation with the higher R-squared. Then the estimated latent variable will be similar to the unrestricted OLS prediction from the better-fitting

---

[4] It can also be shown that at most one of the implied variances can be negative in this case of two indicator variables.

reduced-form equation, perhaps scaled by λ_2 if it is necessary to normalize on the other indicator variable. As a practical matter in estimation, if the reduced-form equations have very different fits by OLS, the iterations will be found to converge faster and more reliably if the model is normalized on the indicator (endogenous) variable with the higher R-squared.[5]

Such practical considerations aside, the question of how to normalize the model is usually seen as a matter of convention and convenience, but it may affect interpretation of the results. In the standard assumption of $\lambda_1 = 1$ for the model of equations (1) and (2), the latent variable is linked to the first-listed indicator variable by the normalization. Reordering the variables will switch another variable to become the normalizing indicator, and hence it will rescale the latent variable. Thus, there is a degree of indeterminacy in scale, which needs to be resolved if the latent variable is to be interpreted as an estimate of the underground economy.

In recognition of this ambiguity, the latent variable is sometimes called an "index". The approach in the literature is to set the absolute level of the estimate by requiring the index to pass through a particular value at a particular time, in a step that is often called "benchmarking" but is more accurately described as calibration. This is analogous to the familiar treatment of an index of prices, in which the series is set in a base period to an arbitrary value of one or 100, and the rest of the series is scaled accordingly. In the present case, the benchmark is not some arbitrary number, but rather is found from other modelling carried out independently of the MIMIC model. If the calibration is multiplicative, it will preserve the proportional relationships in the series (as with a price index). In such a rescaling operation, it will make no difference to the final inference which of the indicator variables is used for normalization. However, as we shall see, the calibration is not always done in this way and, as a result, the inference is not always invariant to the normalization.

THREE APPLICATIONS TO THE UNDERGROUND ECONOMY

I shall present three applications in which MIMIC modelling is used to estimate the underground economy: Giles and Tedds (2002), Bajada and Schneider (2005) and Dell'Anno and Schneider (2003). There is much that is common to these studies, and the latter two papers cite the earlier as a forerunner, but I also find considerable variety among the applications in their approach and interpretation. Unfortunately, the reader is not always informed of these differences by the documentation provided. There are instances in all three works where the description of a procedure, or the context of references to other literature, suggests one approach when, in fact, a different calculation is needed to obtain the stated results. So, while the explanations of *why* something is done are drawn from the papers themselves, I rely on my own careful replications of the calculations to determine *what* is actually done to the data.[6] These replications employ the original data or a close facsimile of them.[7]

---

[5] This is similar to recent findings in the "weak instruments" literature; for example Hamilton, Zha, and Waggoner (2007).
[6] Replication is valuable as a springboard to new inquiry from existing published research, and is an efficient method of purging incorrect results from the body of accumulated knowledge. See McCullough, McGeary, and Harrison (2006) for an evaluation of replication in applied economics and analysis of the data archives of the

This section considers only those aspects of modelling and reporting that are essential to understanding the various ways in which MIMIC modelling is used. I shall focus on the issues of specifying and estimating the model, calibrating the index, and interpreting the resulting time series. Other errors and anomalies uncovered in the process of replicating the three studies are described in the Appendix. This additional information will be useful for readers who seek to understand the substantive results in the three studies.

To simplify the discussion, I shall define a standard notation for the common variables. In Table 1, variables with names of one and two characters appear in at least two of the studies, or are components of constructed variables, while those variables with longer names are used uniquely. I shall use the abbreviated, symbolic names, even when the original study might use a longer description, for example using $\Delta \ln(YD/(P \times N))$ rather than "the proportional growth rate of real, per capita, disposable income". Some fine distinctions may be obscured by this practice (such as the units of measurement or the base year for a price index), but such subtleties can be recovered when they are needed.

Table 1. Definitions of variables

Y	....	nominal observed GDP
C	....	currency held by public
YD	....	nominal disposable income (= $Y - TH - TB + W$)
TH, TB, TI		taxes collected from households and business, and indirect taxes
W	....	welfare state benefits and transfers
P	....	price level
L	....	labour force
M	....	unemployment rate
N	....	national population
S	....	number of self-employed persons
U	....	nominal underground income
$MULT$	....	number of male holders of multiple jobs
$SELF$	....	nominal incomes of self-employed persons
$ERTE$	....	nominal \$Can/\$US exchange rate

All three applications employ two indicator variables (the vector of *y*s) and a short list of causal variables (the vector of *x*s). The indicator and causal variables for each study are listed in Table 2, along with other summary information that will be discussed in detail under the individual studies. The pair of indicators in each case consists of observed GDP in some measure (real, or real per capita, in a logarithmic transformation) and currency holdings by the public in some similar measure. The causal variables are more varied, but typically include a range of tax rates and some measures of real disposable income per capita, the level of employment or unemployment, the extent of self-employment, and welfare state transfers or total government spending. In all three cases there is some sequential differencing of the variables before the model is fitted, as a treatment for unit roots and cointegration, although there are some

differences in the criteria used to make decisions about the differencing. There is also divergence among the applications in the extent to which they standardize the means and standard deviations of the variables before estimation. Further differences will be observed between the three studies in their interpretations of the latent variable and, in particular, in the various ways they calibrate the index after estimation.

Table 2. Summaries of the three studies

Study 1: Giles & Tedds (2002), Model 6
Indicators: $1n(Y/P)^{\dagger}$, C
Causes: *MULT, SELF, YD/(P x L), ERTE, TB/Y, TI/Y, M*
Data: Canada, annual 1976-1995
Specification: in levels
Differencing: levels or first differences or second differences
Undocumented: deviations-from-means and unit standard deviation (standardized)
Index: $100U/Y$ in percentages
Calibration: multiplicatively, to a level of 9.45 per cent in 1986
Base for levels: set by the calibration benchmark.
Study 2: Bajada & Schneider (2005)
Indicators: $1n(Y/(P \times N))$, $1n(C/(P \times N))^{\dagger}$
Causes: $1n(YD/(P \times N))$, $1n(TH/Y)$, $1n(TB/Y)$, $1n(TI/Y)$, $1n(W/YD)$
Data: Australia, quarterly 1966q2 to 2003q3, deseasonalized
Specification: in differences
Differencing: first differences
Undocumented: deviations-from-means
Index: $100\Delta 1n(U/Y)$, integrated and transformed to $100U/Y$ in percentages
Calibration: additively, to a growth rate of 0.0021 per cent in 1980q2
Base for levels: approximately 13.5 per cent in 1968q2?
Study 3: Dell'Anno & Schneider (2003), Model 3-1-2b
Indicators: $1n(Y/P)^{\dagger}$, $1n(C)$
Causes: $(TH + TB + TI)/Y$, G/Y, S/L
Data: Italy, semi-annual, 1960s1 to 2000s2
Specification: in differences
Differencing: first differences, causes and income indicator multiplied by 100
Undocumented: deviations-from-means
Index: $\Delta U/P$, integrated to U/P in units of 10 billion euros
Calibration: none
Base for levels: 19.7 per cent in 1978s2.

† = Normalization on this variable

Study 1: Giles and Tedds (2002)

Before the MIMIC model is estimated, the variables in this study are differenced to the extent that secures their stationarity, according to the results of individual unit root tests. *C* and *SELF* are differenced twice, and most of the other variables are differenced once, while $YD/(P \times L)$ is not differenced at all. The differenced variables are then all transformed into deviations from means and scaled to have unit standard deviation. These last two data operations are not mentioned at all in the published documentation, which is surprising because both are unusual in econometrics. Perhaps these transformations have been made unintentionally, most likely by accidentally invoking an option in the estimation software.[8] Later sections will explore the

---

[8] Tedds and Giles (2005) deny that the variables used in Giles and Tedds (2002) are standardized. However, the estimation results can be replicated if, and only if, the variables are transformed in this way.

consequences for inferences about the underground economy of the (documented) differencing operations and the (undocumented) transformations of location and scale in the variables.

We can interpret Giles and Tedds as specifying the model in the original levels variables, although they estimate the model after variously differencing the variables. This interpretation follows from the way they form the latent variable or index and how they subsequently calibrate the index to become their estimate of the underground economy. In this study, the vector y_t contains the two indicator variables $\ln(Y/P)$ and C, and the vector x_t contains the seven causal variables, *MULT*, *SELF*, etc. We can write the indicators in the estimation model as the vector $\tilde{y}_t$, which contains $\Delta\ln(Y/P)$ and $\Delta^2 C$, after these variables have been transformed to deviation-from-means and scaled to have unit standard deviation. Similarly, we can represent the causes in the estimation model as the vector $\tilde{x}_t$, which contains $\Delta MULT$, $\Delta^2 SELF$, etc., after they have been transformed by location and scale in the same way. Thus, the model is specified just as it is written in equations (1) and (2) with the variables y_t and x_t, but the maximum likelihood estimator is applied after these variables are replaced by $\tilde{y}_t$ and $\tilde{x}_t$. The index in Giles and Tedds, however, is not calculated as $\hat{\eta}_t = \hat{\gamma}'\tilde{x}_t$, which is the direct estimate of the latent variable from the estimation model, but rather as $\hat{\eta}_t = \hat{\gamma}'x_t$, which applies the estimated coefficients to the original, untransformed, causal variables. It is this latter form of index that is scaled in Giles and Tedds' calibration operation, on the grounds that the scale of the index is indeterminate in MIMIC modelling. Clearly, then, they interpret the MIMIC *model* on the original data, even though the *estimates* are derived by fitting the model to transformed data.

Calibration or "benchmarking" in Giles and Tedds is done from a separate currency demand model fitted to data similar to the MIMIC model. From this auxiliary model, an estimate of the underground economy at 9.45 per cent of official GDP is derived for 1986. The index from the MIMIC model is then set to this benchmark, and the rest of the estimated series is found proportionally:

$$ug_t = 9.45 \times \hat{\eta}_t / \hat{\eta}_{1986} \quad \text{for } t = 1976,...,1995 \tag{18}$$

While this formula is not stated explicitly in Giles and Tedds (2002), it is described in words in Giles (1999) and its use by Giles and Tedds is confirmed by replication of their results. It is simply a scaling operation, so it preserves the proportional relationships between the measurements in different years:

$$ug_t / ug_s = \hat{\eta}_t / \hat{\eta}_s \quad \text{for all } t \text{ and } s \tag{19}$$

Thus, the calibrated series will be the same, whichever of the indicator variables is used for normalization, because the arbitrary choice of scale that is imposed by the normalization is removed in the calibration operation.

The scaled series ug_t is interpreted by Giles and Tedds as estimating the underground economy income in Canada as a percentage of observed GDP, that is $100U/Y$. Their resulting estimate is a

20-year time series that grows from a low of 3.46 per cent of GDP in 1976 to a high of 15.64 per cent in 1995, passing through the benchmark value of 9.45 in 1986.[9]

Because of the multiplicative scaling in (18), the overall *level* of this estimate of the underground economy is derived from the benchmark value, which comes from the separate currency demand model. On the other hand, the *time path* of the estimate is due entirely to the MIMIC model. The series is 4.5 times higher at the end of the 20 years than at the beginning, which is equivalent to a compound rate of increase of 7.8 per cent per year. This phenomenal growth rate is more remarkable for being relative to observed GDP, which in real terms grows by 64 per cent in the same period. Thus, according to this estimate, the level of underground income in Canada, in real dollars using the implicit GDP price deflator, increases more than seven times in 20 years. At the same time, the observed economy much less than doubles in size. This astounding growth is the main inference from the MIMIC model.[10]

Study 2: Bajada & Schneider (2005)

Although this study refers to Giles and Tedds as a progenitor, the approach here is very different. Each variable is differenced only once, so the indicators and causes in the estimation model are all quarterly proportional growth rates of the underlying economic variables. This uniform single differencing may be a matter of luck, because the only discussion of the strategy is that "the data used in the MIMIC estimation were differenced after testing for the presence of a unit root" (p.394). However, there is also consistency in the way the variables are uniformly in logarithmic form and are either major economic aggregates measured in real terms per capita, or tax and welfare payments in proportion to an aggregate of income. The variables are all calculated as deviations from means in the estimation model (although that transformation is not documented), but there is no scaling of the variables to have unit standard deviation, as there is in Giles and Tedds.

In this case, it is appropriate to think of the model as being specified *and* estimated in quarterly growth rates. The authors interpret the latent variable in the estimation model as the (percentage) growth rate of the ratio of underground income to observed GDP, $100\Delta\ln(U/Y)$. This quantity is first calibrated, and then integrated up from the growth rates to form an index of the level of $100 U/Y$. A second round of adjustment is employed later to allow the level of the underground economy to be inferred from an estimate of its growth rate.

Again, calibration is done from a currency model estimated from similar data to the MIMIC model.[11] As Bajada and Schneider say, "A quarterly growth rate was chosen from the results of the currency-demand model as a benchmark to produce a growth rate of the underground economy implied by the MIMIC index" (p.394). Although the authors do not specify how this operation is done, from replication of their results it is apparent that the calibration is not the *multiplicative* adjustment of Giles and Tedds, but instead a novel form of *additive* adjustment.

---

[9] The results are shown in Giles and Tedds (2002, Table 7.1 and Figure 7.2).
[10] See Appendix for a further discussion of the modelling and these results.
[11] Calibration in Bajada and Schneider is done from a slightly modified form of the currency demand model of Bajada (1999). See Appendix for a discussion of the currency model and the calibration results.

Suppose we write the latent variable derived by the prediction formula (17) with estimated coefficients as $\hat{\eta}_t = \hat{\gamma}' \tilde{x}_t$. Here, $\tilde{x}_t$ contains the causal variables of the estimation model, in this case $\Delta \ln(YD/(P \times N))$, $\Delta \ln(TH/Y)$, etc., each adjusted to deviations from means. Then, the operation used for calibration by Bajada and Schneider can be written as

$$ugd_t = ugd_0 + \hat{\eta}_t - \hat{\eta}_0 \text{ for } t = 1,...,N \tag{20}$$

Where ugd_0 is the benchmark value of the series of differences, taken from the currency model in the benchmark period $t = 0$, and $\hat{\eta}_0$ is the value of the latent variable from the MIMIC model in the same period. The magnitude of ugd_0 and the timing of the benchmark period are unstated by the authors, but appear to be set at 0.0021 in 1980q2.[12] This procedure simply matches the growth rate from the MIMIC model to that of the currency model in the benchmark period, by adding a constant to the growth rate each period. In contrast to Giles and Tedds, who *scale* the predicted latent variable in the levels model to meet the benchmark, the procedure adopted here is to *slide* the latent variable in the differences model into place against the benchmark.

Bajada and Schneider do not offer any rationale for this additive form of post-estimation adjustment. It certainly does not satisfy the principle emphasised by Giles and Tedds that the scale of the latent variable from a MIMIC model is arbitrary and must be fixed on other information. Here, it is the level of the latent variable that is being adjusted, not its scale. Nor is this form of modelling invariant to the choice of normalizing indicator variable. Normalizing on the income variable rather than currency will change the scale of the coefficients, and hence that of the index. The new scale does not cancel in this procedure, as it does in Giles and Tedds' multiplicative calibration; so in this case, with additive calibration, the choice of indicator for normalization is substantive and not just a mathematical or computational convenience. It might therefore be supposed that Bajada and Schneider attach some meaning to the normalization they adopt. However, all we are told is that "the coefficient on currency holdings is constrained to +1.00 in order to identify the system and make the parameter estimates more easily comparable with one another" (p.393).

It is tempting to suppose that a change of scale in the original variables is equivalent to an additive shift in the logarithms of those variables, but here a constant is added to the growth rates, which becomes an additive linear time trend in the levels of the logarithms, and hence a multiplicative exponential trend in the underlying economic variables when the logarithm transformation is reversed. There is no dimension is which this procedure is an adjustment to fix an unidentified scale. Curiously, however, there is apparently one small virtue of this form of calibration. It turns out not to matter whether the undocumented deviations-from-means transformation of the estimation variables is included or ignored when the latent variable is calculated by the formula $\hat{\eta}_t = \hat{\gamma}' \tilde{x}_t$. The difference between the two approaches will be an additive constant, which will then cancel when the index is adjusted additively to its benchmark.

Bajada and Schneider interpret the calibrated series called ugd_t in (20) as the percentage growth rate of the underground economy relative to official GDP. The growth rate is then integrated to

---

[12] This specification of the benchmark is deduced from inspection of Bajada and Schneider (2005, Figure 4).

get the levels: "Using the currency demand approach to benchmark the starting values of the shadow economy, the MIMIC index was used to generate the level path (as a percentage of GDP) for the shadow economy" (p.395). Unfortunately, no details of this second round of adjustment are provided, and I have been unable to reconstruct precisely the method that was used. In any case, it is misleading to call this second stage calibration, much less benchmarking, because it does not adjust the measuring device against external data; it simply fixes a base point that converts a series of growth rates into a series of levels. Perhaps *anchoring* is a better term.

Taking the results of the second adjustment operation at face value, the level of the underground economy is shown in Table 3 of Bajada and Schneider to hover close to 13.9 per cent of recorded GDP for the period 1993-2003.[13] The annual figure over this decade never moves more than 0.3 of a percentage point from its average. This is a remarkably flat time series by any comparison, both with the estimates for Australia by the method of currency demand modelling in Bajada (1999) and with results reported for other countries. However, since outside information is used to fix both the growth rate of the index (by calibration) and its level (by anchoring), there is not much in this result that can be attributed to the MIMIC model.[14]

Study 3: Dell'Anno & Schneider (2003)

Dell'Anno and Schneider also cite Giles and Tedds as a forerunner, but they employ a methodology that differs in certain crucial ways from both that and the study by Bajada and Schneider. The variables are uniformly first differenced for estimation, apparently without prior testing but rather on the principle that, "In order to eliminate the non-stationarity of the variables, the [causes] are taken as first differences, [while the indicators] are converted in the first differences of logarithm" (p.102, fn.13).[15] Both indicator variables, when differenced, are interpreted as semi-annual growth rates, of real income and nominal currency holdings, respectively. In another parallel to Bajada and Schneider, the causes are taxes or government expenditures in proportion to GDP, or labour force categories in proportion to the total. The variables are all transformed to deviations from means for estimation (again undocumented), although there is no scaling to unit standard deviation.

Also in common with Bajada and Schneider, the model is specified *and* estimated in first differences. However, in this case the authors interpret the latent variable in the estimation model quite differently – as the change in real underground income, $\Delta(U/P)$. This differs in dimension from both indicator variables, which in their differenced form are growth rates of the underlying economic variables. In further contrast to the other studies, the latent variable from the estimation model is not calibrated to an outside estimate, but instead is assumed (implicitly) to be measured in units of 10 billion euros. This quantity is then integrated up from the changes to form an index in the levels of U/P. An external value from other studies is used to fix the overall

---

[13] The final estimates are taken "at face value" because I cannot replicate them. The units of measurement are muddled and there are obvious contradictions between the growth rates in Figure 4 of Bajada and Schneider and the levels in their Table 3. See Appendix for details.

[14] Further implications of fixing both the growth rate and the level of the index are pursued in the Appendix.

[15] There are unit root tests in a sole-authored discussion paper by Dell'Anno (2003), which appears to be an earlier version of the Dell'Anno and Schneider paper.

level of the series to a value of 19.7 per cent in 1978s2. As noted earlier, this is not calibrating the MIMIC index in the manner of Giles and Tedds, but rather anchoring the time path of the levels, which is required because the levels are being constructed from estimates of the changes.

The index is not obtained by Dell'Anno and Schneider directly from the estimation model, but rather is constructed as a separate predictor $\hat{\eta}_t = \hat{\gamma}'\check{x}_t$, where $\check{x}_t$ contains the differenced causal variables. The distinction here is that the deviations-from-means transformation applied to the data to obtain the parameter estimates in the MIMIC model is ignored in forming $\check{x}_t$. There is no additive calibration adjustment here, as there is in Bajada and Schneider, so the two ways of forming the index will differ by a constant. Since this index is being interpreted as the change in real underground income, the constant difference will affect every point in the final series (except the one point where it is anchored on external information).

With no calibration of the latent variable obtained from the MIMIC model, this application does not conform to the principle that the scale of the index is arbitrary and must be fixed on other information. The inferences in this case will depend materially on the choice of indicator variable used for normalization. There are suggestions that the authors are troubled by the contradictions that arise. On the one hand, they recognize that their choice of normalizing indicator $1n(Y/P)$ is material: "this variable ... is chosen as *variable of scale* (or *reference variable*)" (p.105, emphasis in original), and "The choice of the 'sign' of the coefficient of scale (λ_{11}) is based on theoretical and empirical arguments" (p.106).

However, elsewhere they accept that normalization should be a matter of convention and convenience:

> ...in order to estimate not only the relative size of the parameters but their levels, is necessary to fix a scale for the unobserved variable. A natural normalization would be to assign a unit variance to the latent variable but a more convenient alternative is fix a non-zero coefficient to reduced form (p.105, fn.19).

> The value of the fix parameter is arbitrary, but using a positive (or negative) unit value is easier to find out the relative magnitude of the other indicator variables (p.106).

To further confuse the issue of normalization, the authors do not simply choose which of the indicators $1n(Y/P)$ and $1n(C)$ is given a unit coefficient; they specify that $1n(Y/P)$ should have a coefficient of *negative* one. The "theoretical and empirical arguments" for this decision are not made explicit, but it seems that the objective is to ensure that key coefficients in the structural equation for the latent variable have the desired sign. If the normalizing were done in the usual way, the inference would be the unfortunate one that higher growth in the tax burden, or in the size of government, or in the extent of self-employment, all lead to *reductions* in the size of the underground economy.

Given that normalization is arbitrary, in both magnitude and in sign, it is impossible to infer any relationship between the latent variable that represents the underground economy and the endogenous variable on which the normalization is made. However, the authors feel no such inhibition in claiming: "In our analysis, we find evidences to support the hypothesis of negative

relation between Italian shadow economy and official growth rate of GDP" (p.106) and "The relationship between underground economy and growth rate of GDP (Y1) is negative" (p.112).

The final output of this study is a time path of underground income in proportion to official GDP ranging from over 40 per cent in the early 1960s, down to 15 per cent in 1975-77, and then back to around 25 per cent in 2000. Along the way, it passes through the anchored value of 19.7 per cent in 1978 (Dell'Anno & Schneider, 2003, pp.110-111, Figures 2 and 3). The shape of the path depends on the twin assumptions that the index from the MIMIC model is measuring changes in real underground income and the measurement is in units of 10 billion euros. Any other interpretation will give a materially different time path, although both parts of the assumption are quite arbitrary (and unstated in the paper). The choice of the income variable for normalization and the transformation of the variables in the estimation model to deviations from means both influence the result – although the former is arbitrary and the latter undocumented. Furthermore, the assignment of a negative coefficient to the normalizing indicator variable will reverse the sign of the latent variable. Since the latent variable is interpreted as a series of changes, that decision will *invert* the time path of the final result.

DEVIATIONS FROM MEAN AND UNIT STANDARD DEVIATION

I have been able to replicate the estimation results of these three studies without using the specialist LISREL® or Amos™ software. This independent reconstruction of the estimates reveals that the variables have been transformed to deviations from means, and in one case also scaled to have unit standard deviations, although these transformations are undocumented. The finding that all three studies make at least one of these transformations, apparently without the authors being aware of doing so, is at once puzzling and alarming. I shall examine the nature of these transformations and explore their effects on the inference that is made of an underground economy. In every case, the transformation applied in estimation is ignored when the prediction is formed, with the result that the inference acquires some very undesirable properties. I also speculate on how such undocumented transformations might have occurred.

Consider a simple linear regression model between a scalar y and a vector x ($q \times 1$)

$$y_t = \gamma_0 + \gamma_1' x_t + \varepsilon_t \tag{22}$$

where the intercept scalar γ_0 and slopes vector γ_1 ($q \times 1$) are unknown parameters. The error term ε_t is assumed to have zero mean and constant variance, and to be serially uncorrelated for observations $t = 1,\ldots,N$. In a well-known set of results, the least squares estimates are:

$$g_1 = \left[\sum x_t^* (x_t^*)' \right]^{-1} \sum x_t^* y_t^* \text{ and } g_0 = \bar{y} - g_1' \bar{x} \tag{23}$$

Here, the variables x_t^* and y_t^* are transformations of the original variables into deviations from their sample means:

$$x_t^* = x_t - \bar{x} \text{ where } \bar{x} = N^{-1} \sum x_t, \text{ and similarly for } y_t^* \tag{24}$$

The original model can be written as

$$y_t^* = \gamma_1' x_t^* + \varepsilon_t^* \tag{25}$$

which has transformed variables but no intercept. The first equation in (23) indicates that least squares on (25) gives the same slope estimate as the original model (22). The second equation in (23) shows how to extract the implied intercept.

While estimation is the same in both transformed and untransformed variables (provided an intercept is fitted in the latter case), more care is needed when making predictions. For one thing, the models have different dependent variables, so the targets of prediction are different. Using the standard form of the predictor in both cases, $g_1' x_t^*$ predicts y_t^* in the transformed model, while $g_0 + g_1' x_t$ predicts y_t in the original model, in both cases giving an unbiased prediction. However, when a model is fitted to variables that have been transformed to deviations from means, but that transformation is ignored when the predictions are formed, the result will be a hybrid predictor of the form $g_1' x_t$. This makes a biased prediction of *both* y_t^* and y_t. What is more, the bias depends on the intercept in the model, so if any variable in the equation is in logarithmic form, the intercept will change with the units of measurement of that variable, making the whole procedure sensitive to the change in units. This is a clear deficiency in what seems to be common practice in forming the latent variable after MIMIC estimation.[16]

Another transformation that is sometimes considered in linear regression is to write the model with each variable standardized by subtracting its mean and dividing by its standard deviation

$$y_t^{**} = \beta_1' x_t^{**} + \varepsilon_t^* \tag{26}$$

where the standardized variables

$$y_t^{**} = (y_t - \bar{y})/s_y \text{ where } s_y^2 = N^{-1} \sum (y_t - \bar{y})^2 \text{, and similarly for } x_t^{**}$$

have mean of zero and standard deviation of one. The estimates of coefficients in (26) by least squares are called various names; for example, they are "normalized beta coefficients" in Stata™ and "standardized beta coefficients" or just "betas" in SPSS. The connection with the usual estimates is:

$$b_j = (s_x/s_y) g_j \text{ for } j = 1, ..., q \tag{27}$$

Standardized betas are occasionally used to make statements about the relative importance of the independent variables in a multiple regression model. They are invariant to the units in which the variables are measured so, if there is a change of units that rescales one or more of the variables, the standardized betas are unaffected. Again, care is required in making predictions from the transformed model that the transformed predictor variables are used and that the object of

---

[16] The correct predictor in the transformed model, $g_1' x_t^*$, has the property that it is zero on average. If this predictor is interpreted as a series of changes or growth rates and integrated to form an index for the levels, the resulting index has the property that its net change over the estimation period is zero. This will imply that the estimated underground economy is the same size at both ends of the period. None of the three applications actually does this: they either benchmark the differences before integrating or they use a different (and incorrect) prediction formula.

prediction is the transformed dependent variable. Otherwise, as we shall see below, the prediction is not only biased, but is also sensitive to the scale of the units in which the variables are measured.

It is natural in the approach taken by the software packages LISREL® and Amos™ to think of the data being first transformed to deviations from means, and sometimes also transformed to unit standard deviation. The statistical orientation of the user community tends towards multivariate analysis and the use of factor structures to represent patterns of covariance and correlation. The language and assumptions of the software reflect that orientation. Hence, the structural model of Section 2 above might be described as a problem of summarizing the covariances of the data contained in the extended vector $z = (y',x')'$, using a conditional mean with the structure $E(y|x) = \lambda\gamma'x$, a conditional variance $\text{var}(y|x) = \lambda\lambda'\psi+\Theta$, where Θ is diagonal, and without restricting the covariance matrix of x. Given the focus on modelling covariance in this approach, it is often assumed that the means of the variables have already been removed. Hence, the default setting in the software is to subtract the means from each of the variables before fitting the model, thus transforming it in this way. For example, unless there is an "MA" instruction on the "DA" line in the input file, LISREL® will automatically transform the data to deviations from means. This should not surprise the economist: subtracting the means is equivalent to fitting an intercept in a linear regression, which is also the default in most econometric software.

The prior transformation of all variables to have unit standard deviations is also quite natural in this setting. It corresponds with a focus in the analysis on modelling the correlations of the data rather than the covariances. If the model being fitted is one like simple factor analysis that can be described entirely as restrictions on the correlation structure of the data, then it may be convenient to transform in this way. Indeed, LISREL® has options to input the data in the form of a correlation matrix if that is convenient to the researcher. When the data are input as variables, not correlations, there are options that include transforming the variables in the estimation model to have unit standard deviation. Again, in LISREL®, the "SC" option on the "OU" line will give a fully standardized solution. There are equivalent options in other software: for example, in Stata™ the option "beta" on the "regress" command will output the standardized regression coefficients.

As we have noted, a faulty predictor of the latent variable will be employed when the researcher is unaware that the model is estimated on transformed variables. In Giles and Tedds, the estimation variables are fully standardized (transformed to deviations from means and adjusted to unit standard deviation), so the coefficient estimates are invariant to any changes in the units of measurement of the variables. For example, the variable *SELF* is measured in their data file in units of thousands of dollars a year. If all the values of the variable were divided by a thousand or a million, so the new units of measurement become millions of dollars or billions of dollars a year, exactly the same coefficients would be obtained in the MIMIC model because of the standardizing transformation. However, the predictor of the latent variable is formed by applying these standardized coefficients to the original variables. This hybrid form of predictor is not only biased, but is also sensitive to the units in which the variables are measured.

17

In the case of ordinary regression coefficients, any rescaling of a variable is compensated by an inverse scaling of its coefficient, so the product of the two remains invariant when a predictor is formed by linear combination; but when standardized coefficients are applied to non-standardized variables, no such compensation will occur. The coefficient remains constant as the variable is rescaled, so the product of the two elements changes with the scale of the variable. With more than one causal variable in the model, this will not be simply a scaling of the predictor (which might be removed subsequently by multiplicative calibration), but a more complicated set of changes to the relative weights of the variables in the linear combination. Thus, the final inference will be altered materially by the choice of units.

There are further problems with the hybrid predictor as used by Giles and Tedds. When standardized coefficients are applied to variables that are measured on vastly different scales, one or two of the variables are likely to dominate in the linear combination that forms the predictor. In the Giles and Tedds case, it turns out that just one causal variable dominates the latent variable, and hence contributes almost all of the movement over time in the index of the underground economy. That variable is *SELF*, the nominal incomes of self-employed persons, measured in thousands of dollars a year. None of the more plausible variables in their model, such as the various tax rates, has any effect on their estimate.[17]

Bajada and Schneider employ the deviations-from-means but not the unit standard deviation transformation. In principle, the hybrid prediction strategy of applying the coefficients from the transformed model to the original variables will yield a biased predictor in this case. Also, the hybrid predictor will be sensitive to the units of measurement of any of the variables, which are all in logarithmic form. Happily, as we saw earlier, the additive form of calibration they use in forming the index will compensate for the form of the predictor. There remains the issue of a model that is incorrectly described, because the transformation is not reported, nor is the implicit intercept noted. We also observe that the unusual form of calibration in this application imposes an arbitrary solution to the identification problem in the MIMIC model. The results of this study would be substantively different if another, equally arbitrary, normalization of the latent variable were adopted.

In Dell'Anno and Schneider, the data are similarly transformed to deviations from means but not to unit standard deviations. The same criticism applies in this case of a model that is inadequately described, having either undocumented data transformations or a missing intercept parameter. As in the other applications, prediction of the latent variable is biased, because the means of the variables are removed for estimation but included when forming the predictor. Also, the construction of the index is sensitive to the units of measurement of the indicator variables, which both appear in logarithmic form. However, all of these are minor quibbles in the face of the larger problem noted earlier – that the units of the resulting index in this study are simply invented!

---

[17] See Appendix for more detail and references.

DIFFERENCING AND COINTEGRATION

The aggregate time-series data used in all of these studies typically contain trends that may be attributed to unit roots. The reaction in all cases is similar:

> Before one can use the data ... appropriately to estimate models of the form given by [equations (1) and (2) above], one must check for the presence of unit roots. ... [W]e differenced the various data series appropriately to make them stationary. We then used them in this "filtered" form to estimate the MIMIC models... Usually, rather than proceeding directly to modelling after the unit root tests, one would also consider the possibility of cointegration. Unfortunately, there is no established literature to serve as a guide to this procedure in the context of MIMIC models (Giles & Tedds, 2002, p.128).

Dell'Anno and Schneider quote the final two sentences of the above passage and add, "... in some cases, to eliminate the non-stationarity in the time series, the variables are transformed (first differences and growth rates)" (p.107, fn.22). Bajada and Schneider are less informative about their motives and criteria, and simply say "... the data used in the MIMIC estimation were differenced after testing for the presence of a unit root" (p.394).

It is not entirely clear why unit roots are considered to be a problem in this setting. Somewhat earlier in their book than the passage quoted above, Giles and Tedds suggest one issue:

> Essentially, the point is that before one estimates a MIMIC model one must establish the properties of the data; otherwise, the result may be estimates that have undesirable statistical properties and hence measures of the latent variable that are meaningless (Giles & Tedds, 2002, p.104).

A different motive is indicated when these authors later seek to clarify their method:

> It is generally accepted that when modeling with time-series data, these data must first be tested for the presence of unit roots; if these are detected (and in the absence of cointegration), they are rendered stationary in order to avoid the consequences of estimating spurious regressions. That is, the model's coefficients are obtained using the stationary series, but the model's predicted values are calculated using the original data (Tedds & Giles, 2005, p.395).

Thus, two distinct dangers are identified: a meaningless latent variable because the coefficient estimates on which it is formed have undesirable statistical properties, and the risk of estimating spurious relationships.[18]

As I shall show, the act of differencing the variables before fitting the MIMIC model cannot solve the first of these supposed problems, while the second is simply irrelevant to the task at hand. Either the model is a relationship in the levels, in which case differencing is mildly or

---

[18] The term spurious *regression* seems curious here, since the MIMIC model supposedly represents a set of structural relationships, not simply statistical regression.

seriously damaging, or it is a relationship only in the differences, in which case there is no justification for forming an index in the levels. I shall consider both of these possibilities in turn.

On the first hypothesis, consider a model in the original levels of the variables. If the model consisting of equations (1) and (2), together with the assumptions on the variances and covariances of the errors, is a correct description of the process generating the data, there is no reason for concern about unit roots and cointegration. In that case, if the variables have unit roots they must be cointegrated, with two distinct cointegrating vectors since (7) and (8) describe linear combinations of the variables that are stationary (in fact, the linear combinations are white noise). There is a particular relationship between the two cointegrating vectors in this case, which follows from the structure of the MIMIC model. Of course, the conventional asymptotic distribution theory may not apply to the coefficient estimates, because the exogenous (causal) variables x will not have moments that converge in the way that is usually assumed in applications of maximum likelihood to independent data; but the coefficient estimates will be consistent, so the predictor will be cointegrated with each of the endogenous (indicator) variables. As in the standard theory, the latent variable is the fully efficient predictor of the normalizing indicator variable.

Estimating the model after differencing the variables either throws away information relative to fitting the model in the levels of the variables, or it imposes incorrect assumptions on the model. At best, the strategy leads to an efficiency loss, although there may be more serious consequences. On the one hand, provided the coefficients are consistent estimates (they may not be so, see below), the index formed from these estimates and the variables in levels will be cointegrated with the indicator variables. The asymptotic theory indicates that the estimates from the model in levels will be "super consistent" in the sense that they converge to the true parameter values at a much faster rate than the conventional root-N consistency. Thus, the variances of the coefficients in the two approaches may be of different orders of magnitude even in moderately-sized samples. So, while the cost in this case is only inefficiency in the coefficient estimates that arises from needless differencing, such losses may indeed be large. On the other hand, if the coefficients estimated after differencing are not consistent, the latent variable will not be cointegrated with the indicator variables. In that case, the outcome will be a predictor that has no long-run relationship with the endogenous variables it is supposed to predict, which is not a satisfactory solution to the problem of unit roots.

Differencing will return consistent estimates when the model satisfies all the assumptions in the levels variables, provided the *same* degree of differencing is applied throughout. To see this in a single-equation example, suppose the model is

$$y_t = \gamma_1 x_{1t} + \gamma_2 x_{2t} + \varepsilon_t \tag{28}$$

where ε_t is white noise and uncorrelated with x_{js} for $j = 1,2$ and for all s and t. Then, when a differencing operator is passed through the model,

$$\Delta y_t = \gamma_1 \Delta x_{1t} + \gamma_2 \Delta x_{2t} + \Delta \varepsilon_t \tag{29}$$

The error term in the transformed model is serially correlated with a moving average process, but the transformed regressors are still uncorrelated with the transformed errors, so the estimates remain consistent. All that is lost in this case is efficiency.

Now consider what happens with *different* degrees of differencing. Suppose the model in (28) still applies and that y and x_1 are both I(1) and cointegrated, while x_2 is I(0). The strategy described earlier applied to this example amounts to estimating the model

$$\Delta y_t = \gamma_1 \Delta x_{1t} + \gamma_2 x_{2t} + v_t \tag{30}$$

where v_t is just shorthand for the implied error term. By comparing (29) and (30), we see that $v_t = \Delta \varepsilon_t - \gamma_2 x_{2t-1}$. If there is any serial correlation in x_2, the error term in this case will be correlated with one of the regressors. The usual estimation procedure (least squares in this simple illustration) will be inconsistent.

Now, in the converse of the initial assumption, suppose the model does not apply in the levels of the variables but it does apply after the variables have been differenced to stationarity (perhaps with different degrees of differencing in the variables). In this case, the model in the differences will be consistently and efficiently estimated by maximum likelihood. The latent variable will be stationary because it is a linear combination of stationary variables, and it will be a good predictor of the normalizing indicator variable in its differenced form because the assumptions of the model are satisfied. The strategy in two of these studies is to integrate the latent variable from the differences model to become the predictor of the levels form of the normalizing indicator. Now, the latter variable has a unit root (that is why it was differenced) and the integrated latent variable will have a unit root, but there is nothing to connect these two unit roots – the two variables will not be cointegrated. Again, we have the unconvincing setting of an index that has no long-run relationship with the indicator variable that it is supposed to predict.

Giles and Tedds form the predictor by applying the coefficients estimated on the differences to the variables in the original levels. In the special case where the variables are all differenced to the same degree, this method is equivalent to integrating the latent variable from the differences model. In general, then, this method exhibits the problem described in the previous paragraph, that the predictor is not cointegrated with its target. Nor will the problem be ameliorated by different degrees of differencing. Viewed from the perspective of creating the predictor in levels from the estimates on the differences, additional unit roots are introduced when the individual variables are integrated separately and to different degrees. Again, there can be no cointegration between the predictor and its target unless the levels variables are cointegrated at the outset. Additionally, in Giles and Tedds' approach, there is a contradiction between the assumption we noted earlier that the model holds in the original levels of the variables – which implies cointegration – and the apparent need to difference the variables to avoid finding spurious relationships.

The strategy of differencing to stationarity before fitting the MIMIC model pays lip service to the issues of unit roots and cointegration, but lacks any clear purpose. To the extent that the strategy is designed to avoid spurious regressions, that objective would be better served by less reliance on goodness-of-fit criteria (which all three studies report with gusto) and more attention to the logic of the relationships in the model. In any case, the purpose of fitting the MIMIC

model is not to obtain coefficient estimates with standard asymptotic properties, nor to investigate whether significant structural relationships exist, but to condense the information contained in the indicator and causal variables into a time-series index that tracks the unobserved underground economy. That is a prediction question, and needs to be addressed by a strategy for making good predictions.

IS THE MIMIC MODEL APPROPRIATE?

The MIMIC model has its origins in the factor analysis of psychometrics, where the correlations of observable variables are explained by common factors or unobservable latent variables. Whether or not a statistical model is suited to a particular application is to some extent a question of judgment, but there are extensions of the original psychometric factor model where the MIMIC structure seems natural. Suppose the indicator variables are scores on various tests of ability, perhaps differentiated by subject matter such as written and verbal language and mathematics. The unobserved factor influencing all of these outcomes might be called "intelligence". In recognition of its hypothetical origins, it might be agreed to measure intelligence on a scale that for convenience is set to average 100 across the population, with a standard deviation of 15. The causal factors for intelligence will depend on the psychological theory, but they might include various parental and environmental characteristics, such as father's education and mother's nutrition during pregnancy.

This psychometric application to measuring intelligence seems far removed from estimating the underground economy in a MIMIC model. For one thing, the underground economy is not a latent or hypothetical quantity like intelligence; it is all too real, just difficult to measure because the agents who participate in it have every incentive to hide their actions. Unlike the psychometric example where the units of measurement can be resolved by convention, the concept and measurement of income in the underground economy are the same as in the observed economy. Once its scope and units are defined, the level of underground income *is* some number, calculated on a well-defined system of measurement. It cannot be open to the researcher to slide or stretch this calculation to fit whatever scale is found to be convenient. On that ground alone, the MIMIC model seems unsuited to the purpose of measuring the underground economy.

A MIMIC model relates multiple indicators y (p x 1) to multiple causes x (q x 1) through a single latent variable η (scalar). As observed by Jöreskog and Goldberger (1975), two broad implications for the observed variables follow from the assumption of a MIMIC structure. The first is that, apart from scale and some independent measurement errors, the indictors $y_1,\ldots,y_p$ are supposed to be alternative measurements of the *same* thing, namely the unobserved quantity η. The second says that, given the causes $x_1,\ldots,x_q$ and the latent variable η, the indicators $y_1,\ldots,y_p$ are mutually *uncorrelated*. Neither of these properties is convincing in these applications to measure the underground economy.

On the first property, none of the earlier applications makes the argument that the indicator variables in their study are just noisy measurements of the underground economy, up to a scale factor. Indeed, to do so would be ludicrous because of the nature of the variables concerned. The pair of indicators in each case consists of observed GDP in some measure, and currency holdings

by the public in some measure. In Giles and Tedds, the indicators are $1n(Y/P)$ and C while the index is $100U/Y$ in percentage units; in Bajada and Schneider, the indicators are $\Delta 1n(Y/(P \times N))$ and $\Delta 1n(C/(P \times N))$ while the index is $100\Delta 1n(U/Y)$ in percentage units; in Dell'Anno and Schneider, the indicators are $\Delta 1n(Y/P)$ and $\Delta 1n(C)$ while the index is $\Delta U/P$ in units of 10 billion euros. In no case in these free-form interpretations is the index even specified to be in the *same dimension* as the indicator variables, so it is impossible to sustain the idea that the indicators are just scaled and noisy measurements of the latent variable. Even if that problem were somehow fixed, it would still beggar belief to suppose that some function of observed income is an observation *of* the underground economy, just missing an adjustment for scale and clouded by errors of measurement. The same disbelief applies in parallel with the other indicator variable, which is some function of currency holdings. It makes no sense to suppose that some transformed versions of observed GDP and currency holdings are measurements of the *same* unobserved entity, whether or not that entity is called the underground economy. This foolishness is compounded in the examples of Giles and Tedds and Dell'Anno and Schneider by the use of one indicator in *real* income and the other in *nominal* currency.

The second property mentioned by Jöreskog and Goldberger indicates that the dependence structure of a MIMIC model is tightly specified. While the model is usually written in terms of covariances and linear relationships, much clearer statements can be made under the additional assumption that the variables in the model are jointly normally distributed (which assumption is implicit in estimation of the model by maximum likelihood). In particular, the correlation structure in a MIMIC model requires that (i) the indicators y are conditionally independent of the causes x, given the latent variable η, and (ii) the indicators $y_1,\ldots,y_p$ are mutually independent, given the latent variable η. Expressed less formally, these implications say that all of the connections that the indicator variables have with the causal variables, and with each other, are carried through the latent variable.

Both of these implications are unacceptable in the applications being considered here. The first suggests that observed GDP and currency holdings are related to the various causal factors in the model – tax rates, unemployment rates, government expenditures, etc. – only through the size of the underground economy. Such a proposition is inconsistent with every known macroeconomic theory of income determination. The second proposition is equally implausible because it says that currency holdings are unrelated to observed income, once account is taken of the underground economy. If nothing else, that arrangement contradicts the currency demand model used in each of these studies to derive a benchmark value for calibrating the index from the MIMIC model.

CONCLUSIONS

We have explored the use of MIMIC models to estimate the level of underground economic activity. The three applied studies by Giles and Tedds (2002), Bajada and Schneider (2005) and Dell'Anno and Schneider (2003) have been found to be very different, despite their claims to a common parentage. Whether the MIMIC model is related to the simultaneous equations model of the econometrics textbook or the factor analysis of its psychometric origins, it is unconvincing

as a framework for measuring the underground economy. The treatment of unit roots and differencing that makes this a dynamic MIMIC model is also misguided.

The literature applying this model to the underground economy abounds with alarming Procrustean tendencies. Various kinds of sliding and scaling of the results are carried out in the name of "benchmarking", although these operations are not always clearly documented. The data are typically transformed in ways that are not only undeclared but have the unfortunate effect of making the results of the study sensitive to the units in which the variables are measured. The complexity of the estimation procedure, together with its deficient documentation, leave the reader unaware of how the results have been stretched or shortened to fit the bed of prior belief.

The three applications were chosen because the datasets are available to enable replication of the calculations. No other approach would have revealed so clearly what was done to the data to obtain their estimates of underground incomes. There are many other results in circulation for various countries, for which the data cannot be identified and which are given no more documentation than "own calculations by the MIMIC method". Readers are advised to adjust their evaluation of these estimates accordingly.

POSTSCRIPT – A RESPONSE TO THE CRITIQUE

The original manuscript for this paper, dated 2005, had a wide circulation. Given the delay in the paper going to publication, it seems only fair to disclose that, in the meantime, a contrary view has been put. Roberto Dell'Anno and Friedrich Schneider wrote a lengthy and detailed response to the criticisms in this paper. The bottom line of Dell'Anno and Schneider (2006) is: "According with us, the MIMIC model is still one of the best approaches to this purpose." This is not the place for me to engage in any rejoinder – readers who have followed so far will be well equipped to make their own judgements on the merits of the competing analyses and arguments.

REFERENCES

Aigner, D. J., Schneider, F., & Ghosh, D. (1988). Me and my shadow: Estimating the size of the US hidden economy from time series data. In W. A. Barnett, E. R. Berndt, & H. White (Eds), *Dynamic Econometric Modelling* (pp.224-243). Cambridge: Cambridge University Press.

Anderson, R. G., Greene, W. H., McCullough, B. D., & Vinod, H. D. (2008). The role of data/code archives in the future of economic research. *Journal of Economic Methodology*, *15*(1), 99-119.

Bajada, C. (1999). Estimates of the underground economy in Australia. *Economic Record*, *75*(231), 369-384.

Bajada, C., & Schneider, F. (2005). The shadow economies of the Asia-Pacific. *Pacific Economic Review*, *10*(3), 379-401.

Breusch, T. (2005a). The Canadian underground economy: An examination of Giles and Tedds. *Canadian Tax Journal*, *53*(2), 367-391.

Breusch, T. (2005b). Australia's cash economy: Are the estimates credible? *Economic Record*, *81*(255), 394-403.

Dell'Anno, R. (2003). *Estimating the Shadow Economy in Italy: A Structural Equation Approach*. Working Paper 2003-07, Institut for Økonomi, Aarhus University. Retrieved from ftp://ftp.econ.au.dk/afn/wp/03/wp03_07.pdf.

Dell'Anno, R., & Schneider, F. (2003). The shadow economy of Italy and other OECD countries: What do we know? *Journal of Public Finance and Public Choice*, *21*(2-3), 97-120.

Dell'Anno, R., & Schneider, F. (2006). *Estimating the Underground Economy by using MIMIC Models: A Response to T. Breusch's Critique*. Working Paper No. 067, Department of Economics, Johannes Kepler University of Linz. Retrieved from https://ideas.repec.org/p/jku/econwp/2006_07.html.

Frey, B. S., & Weck-Hannemann, H. (1984). The hidden economy as an "unobservable" variable. *European Economic Review*, *26*, 33-53.

Giles, D. E. A. (1999). Modelling the hidden economy and the tax-gap in New Zealand. *Empirical Economics*, *24*, 621-640.

Giles, D. E. A., & Tedds, L. M. (2002). *Taxes and the Canadian Underground Economy*. Toronto: Canadian Tax Foundation.

Goldberger, A. S. (1972). *Structural Equation Methods in the Social Sciences*. Amsterdam: North-Holland.

Hamilton, J. D., Zha, T., & Waggoner, D. F. (2007). Normalization in econometrics. *Econometric Reviews*, *26*(2-4), 221-252.

Helberger, C., & Knepel, H. (1988). How big is the shadow economy? A re-analysis of the unobserved-variable approach of B.S. Frey and H. Weck-Hannemann. *European Economic Review*, *32*, 965-976.

Hill, R. (2002). The underground economy in Canada: Boom or bust? *Canadian Tax Journal*, *50*(5), 1641-1654.

Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*(351), 631-639.

McCullough, B. D., McGeary, K. A., & Harrison, T. D. (2006). Lessons from the JMCB Archive. *Journal of Money, Credit and Banking*, *38*(4), 1093-1107.

Pagan, A. (1979). Some consequences of viewing LIML as an iterated Aitkin estimator. *Economics Letters*, 3, 369-372.

Schneider, F., & Enste, D. (2000). *Shadow Economies around the World: Size, Causes, and Consequences?* IMF Working Paper No.26, International Monetary Fund, Washington, DC.

Smith, R. S. (2002). The underground economy: Guidance for policy makers? *Canadian Tax Journal*, *50*(5), 1655-1661.

Tedds, L. M., & Giles, D. E. A. (2005). Response to Breusch's critique. *Canadian Tax Journal*, *53*(2), 392-395.

Zellner, A. (1970). Estimation of regression relationships containing unobservable variables. *International Economic Review*, *11*, 441-454.

APPENDIX – FURTHER PROBLEMS IN THE THREE APPLICATIONS

Negative variances

A difficulty that arises in all three of the studies described is that the estimates are *inadmissible*, in the sense that one of the variance estimates is negative. This outcome is obtained whether the model is estimated by the LIML/GLS procedure or by the packaged solution in LISREL®. In both Giles and Tedds (2002) and Dell'Anno and Schneider (2003), the problem parameter is ψ, which represents the variance of the latent variable. The LISREL® output file in these cases includes the prominent message "WARNING: PSI is not positive definite". It seems that the MIMIC model is *not* a good description of the data in either of these applications, despite the many measures of goodness of fit and the extensive diagnostic testing reported with the estimation results.

In Bajada and Schneider (2005), the offending parameter is θ_1, the variance of the measurement error on the first indicator variable. Again, the solution for a variance is negative, so the estimated MIMIC model is inadmissible, despite being an apparently good fit. In this case, LISREL® does not signal the problem quite so clearly, since no warning message is given. The problem is further obscured by the poor choice of units of measurement for the indicator variables. Both indicator variables in this study are quarterly *proportional* growth rates of macroeconomic variables (real per-capita income and currency holdings). These are quite small numbers, with at least one, and often two or more, leading zeros after the decimal point. The *variances* of such small numbers will be an order of magnitude smaller, because of the squaring operation in forming a variance. More than that, the parameter is the variance of the *observation error* in the variable, which will be that much smaller again. Therefore, these parameters have values that will not be readable in an output field that provides for a moderate but fixed number of decimal places, and they will be completely invisible in the default-width field of two fixed decimal places printed by LISREL®. The answer a researcher will see for each variance estimate in this case is zero. The only signal that something is wrong with the estimate of θ_1 is the negative t-ratio given for this parameter.

Other problems

Giles and Tedds (2002)

We have already noted that the (undocumented) use of standardized variables in Giles and Tedds' estimation model, together with the original variables in the prediction formula, makes the whole procedure sensitive to the units of measurement. As a complication of this sensitivity, their estimate for Canada has nothing to do with most of the causal factors in their model. As shown in Breusch (2005a), their index is almost entirely a rescaling of the variable *SELF*, which is an economy-wide aggregate measured in nominal Canadian dollars. Thus, the major part of the astounding growth they report in the underground economy over 20 years is due to inflation in the price level, while a lesser part is due to expansion of the real size of the Canadian economy, and even less to the composition of the real economy. Nothing of their estimate can be associated with the more plausible factors that they list among their causal variables, such as the number of self-employed persons relative to the rest of the labour force, or the various tax rates.

Their estimated growth rate is not even approximately a measure of the underground economy in Canada.

Also in Breusch (2005a), it is shown that the key parameters are unidentified in the currency demand model used by Giles and Tedds to calibrate the series, so the overall level of the series is not really an estimate at all, but rather a numerical accident. Vastly different "estimates" can be obtained by innocently tweaking some features of the method that should be irrelevant, such as the starting values for the nonlinear algorithm or the software package used for estimation. As with the growth rate, the level of their reported series has nothing to do with measuring the underground economy.

One additional problem in Giles and Tedds – which, in the context of the other problems, is of interest only to researchers seeking to replicate their results – occurs where a variable is not actually differenced as stated. The unemployment rate variable M is described as I(1), and it is reported that all integrated variables are differenced to stationarity. In contradiction to this statement, their MIMIC estimates and subsequent calculations can be replicated only if M is not differenced.

Bajada and Schneider (2005)

The vague language and skimpy reporting of the procedures in this paper frequently make it difficult to tell *what* is being calculated. There are also confusing lapses in accuracy. For instance, the quantities plotted in their Figure 4 are called "growth rates", and the vertical axis is labelled "%", although both of these attributes are likely to be wrong. My replication of their calculations suggests that the values plotted at an annual frequency are not annual rates of growth, as a reader might expect, but instead *quarterly* growth rates that have been averaged over the four quarters of the Australian financial year. The interpretation of the latent variable from the fitted MIMIC model as a *percentage* growth rate seems unwarranted, too, since all of the variables in the model are proportional, not percentage growth rates. Taken together, these corrections suggest that the numbers in Figure 4 should probably be multiplied by 400.

Another confusion revealed by replication of the results is a reversed set of labels in the legend of Figure 4, so what is called the "Currency-demand" line is actually the "MIMIC" result, and vice versa.

The outcomes of the calibration and integration operations are only partially revealed in the paper. In particular, the interim inference about the growth rates in Figure 4 covers only the period from 1980 to 2003, while the final inference about the levels in Table 3 is restricted to an even smaller range from 1993 onwards. There are apparent errors even in this subset of the results, where the growth rates and the levels are mutually inconsistent. To sidestep the problem of the reversed labels in the legend of Figure 4, we can consider only periods in which the currency and MIMIC methods agree on the direction of change. However, there are instances such as the period 1993-2000, in which all of the growth rates are said to be positive, yet in Table 3 for this period there are *falls* in the levels by both methods.

Calibration in this paper is done from a slightly modified form of Bajada's (1999) currency demand model. The difference here is that the excess sensitivity measures of taxes and welfare benefits are expressed in real per-capita terms rather than percentages of GDP. Breusch (2005b)

shows that the original Bajada method is highly sensitive to the units of measurement. In particular, changing the measurement of tax payments from a percentage to a proportion of GDP produces a very different inference about the underground economy (in fact, the estimates become negative!). Exactly the same objection applies in this case, where the substantive results will change when some other units of measurement are used. Replication shows that the results of the paper require the excess sensitivity variables, tax and welfare benefits, to be measured in single dollars per capita, with a 2001-02 price base. Any other scale will give a different outcome. As an example, if the variables are measured in units of thousands of dollars per capita, the results become nonsensical: the "underground economy index" of Figure 2 plummets over time until it is approximately –0.5 by the end of the period.

A second problem with the Bajada method is that the value of income velocity – to which the estimates of underground incomes are directly proportional – is set many times too high. An assumption is made that the income velocity of currency in the underground economy is equal to the ratio of income to currency in the observed economy. While this may have some superficial appeal, it ignores the very small part that currency represents in the money supply of the observed economy (currency is well under 10 per cent of M3 in Australia). Hence, the work that currency does in the generation of observed incomes is vastly overstated by this assumption. Setting the ratios of income to currency in the two sectors to be equal then transmits this exaggerated role of currency to the estimates of underground incomes. Much of the literature using currency modelling to estimate the underground economy makes a similar-looking assumption, but in these cases it is equality across sectors in the income velocity of *total money supply*. While there is some variation in this literature because of different definitions of money, the values of velocity are from one-fifth to one-fifteenth of the value assumed by Bajada. The estimates of incomes in the underground economy in Bajada and Schneider can be reduced in the same proportion.

There is an interesting claim in the paper that finding "very similar results" between currency and MIMIC models somehow validates both forms of modelling (pp.395-396). Given the two-stage processes of calibration and anchoring, as described above, it is clear that their MIMIC results have been tied directly to those of the currency model. Both the level and the rate of growth of the underground economy in the MIMIC results are fixed to the currency model. Then, with the very small estimated rates of growth, the estimates of the levels in either case hardly move from their benchmark value, so it is no surprise that the two sets of results are similar for long periods, because the results called "MIMIC" are almost entirely drawn from the currency model. Furthermore, the similarity or otherwise of the results from the two models is hard to judge when we are shown the final outcome for only eleven of the thirty-seven years of data that are available.

Dell'Anno and Schneider (2003)

There are some small errors and inconsistencies in this paper, which become apparent on replication of the results. In particular, if the variables for tax burden, real government consumption and the rate of self-employment are percentages, as defined in the text of that paper, they should be similarly described in Appendix 1. The variables are then to be multiplied by 100. If these variables are indeed percentages, then the published coefficients indicate that the

other causal variables in the preferred model are also in percentage form. The income indicator variable also needs to be multiplied by 100 to make it a percentage, but scaling of the currency variable is uncertain because there are insufficient decimal places in the published coefficient to see anything but leading zeros! Most likely, this variable is a ratio not a percentage.

The variable described in the text as "real government consumption (in percent of GDP)" is in fact G/Y, and thus the ratio of the two nominal variables. It is not G/P, as reported in Appendix 1, nor is it a more complicated variable involving multiple price indices, as might be inferred from the description in the text. The data period for estimation is unstated in the paper, but the results are most closely replicated by using 1960s1 to 2000s2 (although effectively the data begin in 1962s2 due to the creation of lags and missing observations in the currency variable).

An external estimate is used to anchor the series, so that the growth rates from the MIMIC model are converted into a time series of the level of the underground economy as a percentage of recorded GDP. The overall level of the final product of Dell'Anno and Schneider is due entirely to this external estimate, since only the variations up and down from the anchor point come from the MIMIC model. The anchor value of 19.7 per cent in 1978s2 is obtained as the simple average of five other estimates by various methods (one of which is itself the average of two others). Most of these prior estimates come from an unpublished working paper by Schneider and Enste (2000), in which they are documented as "own calculations".