Measuring the unobservable: estimating informal economy by a structural equation modeling approach

Roberto Dell'Anno[1]

Abstract

This article proposes a new approach to estimate the informal economy (IE) by using Structural Equation Modeling (SEM). Using a Monte Carlo Simulation and empirical analysis of the Italian IE as an example, we provide general conclusions on the reliability and limitations of the SEM approach to estimate the IE. Practical guidelines on how to apply this method and the way to deal with the most problematic issues are provided. We conclude that the SEM approach may be effectively used, as a complementary method to the National Accounting Approach, to adjust official statistics for the presence of the IE.

1 Introduction

Measuring the Informal Economy (IE) is important for economics because statistics on the IE allow for decomposing economic growth into formal and informal sources, which may be important when formulating counter-cyclical and structural policies (Quiros-Romero et al., 2021). IE influences economic performance through several channels and has relevant repercussions on many aspects of the economic and social life of a country. On the one hand, it reduces efficiency in goods and labor markets, worsens economic and social institutions, decreases tax revenue, reduces potential worthwhile public expenditure on infrastructures, education, research, health. On the other hand, it creates an extra value-added that can be spent on the official economy and, mainly for less developed economies or countries with a high

✉ Roberto Dell'Anno
rdellanno@unisa.it

1   Department of Economics and Statistics and CELPE, University of Salerno, Via Giovanni Paolo
     II, 132, 84084 Fisciano (Sa), Italy

unemployment rate, it often represents a social buffer and a source of job opportunities for low-skilled workers and poor. The size of the IE also affects society indirectly. For instance, de Soto (1989) sees informal activities as efficient market forces that emerge as a reaction to over-regulation and government oppression (Biles, 2009). For this reason, IE may be seen as a "signal" for policymakers of overburden in regulation, taxation, and other government-induced distortions. According to these multifaceted effects of informality on economic growth and income distribution, knowledge of the size of the IE is not only an econometrician task but an important topic for economics.

In general, it is possible to identify three main topics in the literature on the IE[1]: a definitional issue (i.e. what is "informal"), a measurement issue (i.e. "how" to estimate its size), and a theoretical explanation of the "informality". In this paper, we focus on the measurement issue and, specifically, on one of the most applied and controversial estimation methods: the Multiple Indicators and Multiple Causes (MIMIC) or Model approach.

The OECD (2002) distinguishes two principal approaches to estimate the Non-Observed Economy (NOE)[2]: the econometric methods and the National Accounting Approach (NA) to estimate the NOE. The former can be furtherly split into three strands. The "direct methods" are based on direct information regarding informal activities collected by (a) tax returns or administrative data; (b) questionnaire surveys; (c) experimental data. The "indirect methods" estimate the size of the IE by measuring the "traces" that it leaves in official macroeconomic data. This strategy includes: (a) the discrepancies methods (e.g. discrepancy between national expenditure and income statistics, or between the official and actual statistics of the labor force); (b) monetary methods, e.g. currency demand approach, physical input method. The third method is based on the statistical approach of Structural Equation Modeling (SEM) and follows the pioneering work of Frey and Weck-Hanneman (1984). It consists in applying a particular model specification of the SEM: the MIMIC model. According to this method, the IE is considered as an 'unobserved' (latent) variable that affects multiple observable indicators and is affected by multiple observable causes.

First of all, it is important to clear the field from misunderstandings and futile competitions between NA and "econometric" approaches on the reliability of the estimates of the IE. It is unambiguous that the NA is the most reliable method to measure the IE. We believe that quality data lead to quality inferences and that no statistical method turns bad data into good data. National statistical offices have more resources, extensive and better data (e.g. microdata collected by ad-hoc surveys, exclusive access to administrative and confidential data collected by public administrations) than the scholars who usually apply econometric approaches. Due to this, the estimates of the NOE provided by statistical offices are the best source to know the size of the IE. Nevertheless, there are some shortcomings and

---

[1] See Schneider and Enste (2000), Slemrod and Weber (2012), Dell'Anno (2016, 2021) and Ulyssea (2020) for a survey on definition(s), theoretical and empirical approaches to informality.

[2] This aggregate includes the usual definition of the IE that is labelled as underground economy.

limits that still suggest to estimate, and analyze, the IE also by using the econometric approaches: (1) the estimates based on the NA are usually published with a significant time lag and short time dimension; (2) the national offices of statistics have different resources (e.g. know-how, data available) and attention (e.g. due to political pressure, institutional factors) to estimate the NOE, and this affects both the degree of coverage of the different types of the NOE and the reliability of the estimates across countries and over time. Accordingly, estimates of the IE based on NA are often internationally and intertemporally incomparable; (3) Disaggregated data on the IE are usually unavailable for scholars outside national institutes of statistics; (4) the process of estimation applied to adjust official statistics for the presence of the NOE are not submitted to external peer review system and not replicable. It means that, according to the standard method for research validation, NOE estimates are in some way nontransparent. These shortcomings make NA estimates problematic for economic purposes and cannot be considered as a perfect substitute for econometric estimates of the IE. For that reason, exploring the other estimation methods should be considered as a valuable aim for economic research.

As we explain in the following sections, the NA and the MIMIC "econometric" approach can be used as complementary methodologies to predict IE. Specifically, on the one hand, the MIMIC approach requires calibrating the estimates by (at least two) exogenous values of the IE and the best candidates for this are the NA estimates of the Underground Economy. On the other hand, NA may benefit from the MIMIC method to predict the IE, in some sectors or periods where NA estimates are not available.

This paper makes a number of contributions. First, we present the MIMIC model as a special case of a SEM and introduce in this literature the Partial Least Square (PLS) estimator of SEM as an alternative algorithm to the standard (i.e. Covariance-Based estimator, hereafter CB) approach to estimate a MIMIC model. Second, given that one of the critiques against the MIMIC approach is related to the lack of transparency in exposition and replicability (e.g. Breusch, 2016; Feige, 2016), we describe, step by step, how to estimate the IE and specify a CB-MIMIC model in order to apply the PLS estimator. Third, we propose three (complementary) methods to convert the latent variable scores into a time-series with a "real" unit of measure (e.g. IE as a share of official GDP or as monetary value). Fourth, given that the most important constraint to assess methods to estimate IE is due to the impossibility to compare predicted and actual (non-observable) values of the IE, we elude this obstacle, by assessing the reliability of the MIMIC approach by a Monte Carlo (MC) simulation. The MC analysis has been set to simulate "realistic" values and, by following Andrew et al. (2020) suggestions to improve transparency in structural research, we focus on the issue of the identification, in robustness of results under several alternative assumptions and under which circumstances SEM approach reverse the predicted trend of the IE. For these reasons, our findings may provide practical guidelines to apply the SEM approach in this field of research. Fifth, we test if MC findings hold also in analyses of real-world phenomena, namely, by estimating the Italian IE from 1995 to 2020. The paper ends with some general conclusions on the reliability and limitations of the SEM approach.

2 MC analysis of the SEM approach to estimate IE

In general, the empirical literature on methods to estimate IE is a divisive topic. In particular, since the first estimates based on the MIMIC approach have been published by Frey and Weck-Hannemann (1984), the MIMIC approach has encountered conflicting reactions among scholars. On the one hand, this methodology allows wide flexibility, therefore it is potentially inclusive of all the indirect methods and thus theoretically, superior to others. For instance, Thomas (1992) states that the only real constraint of this approach is not in its conceptual structure but the chosen variables. Cassar (2001) emphasizes as, in contrast with other indirect methods, it does not need restrictive or implausible assumptions to operate (with exception of the "calibrating value"). Zhou and Oostendorp (2014), by using firm-level data, show as the MIMIC provides a relatively more accurate estimate of underreporting than the direct and indirect approaches. On the other hand, scholars (e.g. Breusch, 2016; Feige, 2016; Kirchgässner, 2016; Slemrod & Weber, 2012) question the reliability of MIMIC predictions.

One of the reasons that make the debate on methods to estimate IE particularly heated is that, differently from other topics of applied economics, there is not an "acid test", where the "observed" values of the IE are compared to the "predicted" ones. This leaves an aura of uncertainty on the predictive performances of the estimation approaches.

Although already existing studies compare CB-SEM and PLS-SEM estimators by MC analysis (e.g. Goodhue et al., 2012a, 2012b; Sarstedt et al. 2016), this article is the first to check the accuracy of predicted values of the latent variable (so-called latent variable scores) to the "unobservable" construct by simulating "realistic" values of the IE. The MC analysis consists of 2 steps. The first step deals with the theoretical and statistical background of the MIMIC approach. The first sub-section describes how the "true" values of the IE and pseudo-random observed variables are simulated. The second subsection presents the standard MIMIC model (i.e. model A) and explains how can be re-specified to apply the PLS algorithm. Moreover, we show a SEM specification that includes relationships between "causes" and "indicators" variables (i.e. the MIMIC B). The second step (Sect. 2.2) explains the identification problem in the MIMIC and the consequent importance of calibration. Furthermore, it proposes three calibration methods to assign a unit of measure to the latent scores.

2.1 First phase: MIMIC specifications

2.1.1 The generation of variables

According to the MIMIC approach, the IE is a "latent" variable that is both affected by a set of observed variables (the so-called structural model) and affects other observable indicators (i.e. the measurement model). The "true" models (i.e. number and characteristics of observed variables in structural and measurement model)

and "true" coefficients (hereinafter labeled with the superscript "*tr*") generate simulated values of IE which closely resemble commonly encountered data in this type of literature. Precisely, the IE is assumed to depend on a linear combination of five observable causes:

$$\text{Inf}^{tr} = a_0^{tr} + \sum_{i=1}^{5} a_i^{tr} x_i + \zeta \tag{1}$$

As the measurement model concerns, the three indicators (y_j) of the IE are generated as follows:

$$y_1 = b_{01}^{tr} + b_1^{tr}\text{Inf}^{tr} + c_4^{tr}x_6 + \varepsilon_1 \tag{2}$$

$$y_2 = b_{02}^{tr} + b_2^{tr}\text{Inf}^{tr} + c_1^{tr}x_5 + \varepsilon_2 \tag{3}$$

$$y_3 = b_{03}^{tr} + b_3^{tr}\text{Inf}^{tr} + c_2^{tr}x_4 + c_3^{tr}x_5 + \varepsilon_3 \tag{4}$$

where the "causes" (x_i) are pseudo-random variables distributed according to normal (N) or uniform (U) distribution;[3] the "indicators"[4] (y_j) depend on the *Inf*^{tr}; the direct effect of some "causes" (e.g. x_4, x_5) or a control variable (e.g. x_6) on the indicators; the ζ and ε_j are (pseudo-random) error terms, normally distributed with mean zero and unit variance. We specify two MIMIC models, to take into account the difference in the goodness of fit between the standard MIMIC specification used in literature i.e. the restricted version (labeled as model A), and a more complex one (i.e. unrestricted model) labeled as model B. For the reduced model A, the parameters constrained to be equal to zero are the direct effects of some "causes" of IE on the indicators y_j (i.e. c_1 =c_2=c_3=0) and the control variable of the "reference indicator" (y_1) (i.e. c_4=0).

2.1.2 The MIMIC model specification

The basic intuition of the MIMIC model is that IE is an endogenous (i.e. related to a set of variables that explain them) latent construct ($\eta_1 = Inf$). Following the usual conventions to graph SEM models,[5] the MIMIC

---

[3] To increase "realism", we assume that all variables are expressed in percentage points and range over values related to some of the most common determinants of IE. Specifically, we assume that x_1 may be an index of tax burden (20–30), x_2 an index of flexibility of labor market (15–25); x_3 self-employment rate (5–25), x_4 an index of Labor cost (1–20), x_5 a proxy of institutional quality such as Rule of Law or Economic Freedom (1–10).

[4] The indicators may be proxies of the IE based on indirect methods to estimate informality, proxies of tax evasion or undeclared work or whatever measurable variables highly correlated to the size of the IE.

[5] Specifically, we indicate observed variables as rectangles, the latent variables and errors terms by circles or ovals, paths or loadings (or regression effects) that connect variables or errors terms as single-headed arrows and double-headed arrows for covariances. The covariances among *x* are constrained equal to zero in the MC analysis to preserve comparison to the OLS estimates. This assumption will be removed in the empirical analysis of the Italian informal economy.

representation of the Eqs. (1–4) can be described by the path diagram of Fig. 1.[6]

Figure 1, without coefficients γ_i for $i=6,…,9$, is the standard specification that is applied in this literature (i.e. Model A). By including the covariates variables in the measurement equations, we get the unconstrained MIMIC model (i.e. Model B) where unconstrained paths and observed variables are shown with dot-line.

According to the Model B specification, the coefficients γ_i are expected to be equal to the coefficients a of Eq. (1) when $i=1,..,5$, to the coefficients c (Eqs. 3 and 4) when $i=6,7,8$, and to the coefficient b_4 of Eq. 2 when $i=9$. Moreover the coefficients λ_j^y should be equal to the values b (Eqs. 2 and 4).

To estimate Eqs. 1–4 by a PLS-MIMIC model, we need to constrain some SEM parameters of the CB-SEM specification.[7] As a result, the output of these two SEM estimators can be compared to analyze their performances to predict the IE and to estimate the "true" coefficients.

Figure 2 shows a "full-SEM" specification corresponding to the MIMIC Model A. It includes five exogenous latent concepts (ξ_i) that are fixed equal to the "causes" (x_i) by appropriate parameters constraints.[8] The advantage of the "full-SEM" specification is that, in this form, model A can be estimated by the PLS algorithm.

As Model B concerns, differently from Model A, it cannot be successfully estimated by the PLS algorithm because there is a manifest variable (i.e. x_6) affecting one of the (reflective) indicators of the latent construct (i.e. $c_4^{tr} \neq 0$ in Eq. 2).[9] Online Appendix A provides the formal definition of the Maximum Likelihood (ML) estimator, which are the constraints on parameters required to shift from a MIMIC B specification to a full-CB-SEM form, and explaining the differences between CB and PLS estimators of a SEM.

As the PLS estimator concerns, although the literature shows mixed opinions on its predictive performances, scholars agree that one of the reasons for the growing success is that, differently from CB-SEM, the PLS algorithm always converges. This characteristic is particularly valuable in SEM applications aimed to estimate IE because some of the most common sources of non-convergence of the CB algorithm, such as small sample size and measurement model including less than three relevant indicators, are the most common constraints encountered by scholars in this type of applications.

---

[6] Given that we move from 4 single equations to a system of equations, we need to specify additional hypothesis on the relationships among equations. We assume that the covariance matrix of the x and the matrix among the covariances of measurement errors ε are diagonal in order to be consistent with the Eqs. 1–4. Both the hypotheses may be modified.

[7] This procedure is due to the impossibility to estimate by PLS-SEM a model where the latent construct is defined simultaneously by reflective and formative measurement models.

[8] In particular, we constrain the variances of the errors terms $Var(\delta_i) = 0$ and the measurement coefficients $\lambda_i^x = 1$. An alternative CB-equivalent MIMIC model may be obtained by generating three endogenous latent variables, one for each indicator, instead of six exogenous latent variables, one for each x. As a consequence, the MIMIC model A (Fig. 1) and this "full-SEM" specification (Fig. 2) generate the same estimates.

[9] Given that in PLS-SEM each latent construct has to be "measured" at least from one manifest variable, the PLS estimates of the IE corresponds to the standardized values of its indicator, i.e. $\hat{\eta}_{1,t} = (y_{1,t} - \mu_{y_1})/\sigma_{y_1}$.
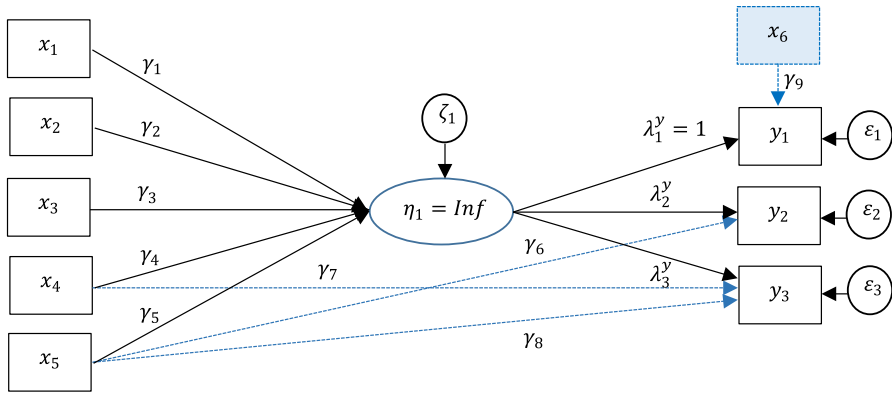
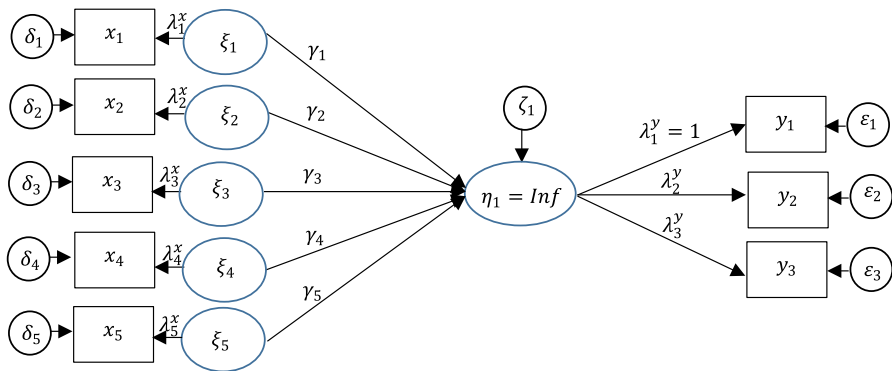**Fig. 1** Path Diagram of MIMIC A and B (CB-SEM)

Fig. 2 MIMIC A in Full-CB-SEM form (as estimated by PLS)

2.2 Second phase: identification and calibration of the MIMIC model

The second phase of analysis deals with the most controversial issue of the MIMIC approach, i.e. how can SEM estimates be converted into actual values of IE? Before introducing the calibration methods, it is important to outline that the calibration methods aim to solve the issue of indeterminacy of the estimated parameters in the SEM (that is due to the identification issue).

This issue, and a possible solution, was originally outlined in the first paper that introduced the MIMIC model in the literature (i.e. Jöreskog & Goldberger, 1975). They observed that the matrixes of structural and measurement coefficients remain unchanged when measurement coefficients are multiplied by scalar and structural coefficients and standard deviation of the structural error (ζ) are divided by the same scalar. The "creators" of the CB-SEM suggested removing this indeterminacy in the structural parameters, by normalizing (hereinafter "anchoring") the variance of the latent factor to one. However, an alternative anchoring method, and

more commonly used in applications of the MIMIC to the IE, is to fix equal to unit a measurement coefficient of the latent variable (usually, the indicator expected to have the highest correlation to the IE).[10] These two anchoring methods (i.e. constraining the variance of the structural equation to one—the $Var(\zeta_1) = \psi_1 = 1$—or fixing equal to unit the first loading factor of measurement model—$\lambda_1^y = 1$) lead to estimate structural and measurement coefficients linked to the latent factor (i.e. $\hat{\gamma}_i$ with $i = 1,\ldots,5$ and λ_j^y) that differ from the "true" value by a scale factor. In this research, we opt for the first anchoring method ($\hat{\lambda}_1^y = 1$), because it is more common in this literature, but the results of this article may be replicated also by applying Jöreskog and Goldberger's (1975) anchoring approach.[11]

In conclusion, both the anchoring approaches do not "solve" the indeterminacy of MIMIC estimates—because all the absolute values of parameters are still arbitrary—but they make it possible to estimate the SEM by fixing a (unknown) scale for all the parameters of the model. That is to say that, due to identification issue, in SEM there is an infinite number of different sets of coefficients (and latent scores) that will fit the model equally well as possible because parameters are a scaled version of true parameters.

As the PLS estimator concerns, also in this case the metric of the latent variable is indeterminate. PLS algorithm standardizes, to a mean of zero and a variance of one, latent scores in each iteration therefore, after convergence, the scores and the path coefficients have to be rescaled to provide a variable in a real metric. An intuitive calibration may consist of replacing the zero mean and unit variance of the latent construct with an exogenous estimate of the mean and variance of the IE.[12]

---

[10] This anchoring method has been preferred because it has a more intuitive meaning, i.e. each unit in the latent variable corresponds to a unit of change in the indicator. In SEM jargon, manifest variable linked to the latent construct by the fixed coefficient is usually defined as "reference" indicator and the measurement coefficient is labeled as "coefficient of scale" (e.g., $\lambda_1^y = 1$).

[11] Given that this issue is often a source of misunderstandings among scholars, an example of the mathematical relationship between MIMIC coefficients based on different anchoring methods may be helpful. Let's assume to identify the MIMIC by anchoring the latent variable to the reference indicator (*ri*), i.e. fixing $\lambda_1^{ri} = 1$, and label these estimated (structural and measurement) coefficients as $\hat{\gamma}_i^{ri}, \hat{\lambda}_i^{ri}$. Let's assume now to identify the MIMIC by the alternative approach proposed by Jöreskog and Goldberger's (1975), i.e. $Var(\zeta_1) = 1$. The structural coefficients, in this second case, are proportional to the estimates obtained in the first anchoring method according to the following relationship: $\hat{\gamma}_i^{ve} = \hat{\gamma}_i^{ri}/(\hat{\psi}_1^{ri})^{0.5}$ for $i = 1,..,5$ and the measurement coefficients according to $\hat{\lambda}_j^{ve} = \hat{\lambda}_j^{ri}(\hat{\psi}_1^{ri})^{0.5}$. Likewise, we can convert estimated coefficients from the second anchoring method ($\hat{\gamma}_i^{ve}, \hat{\lambda}_j^{ve}$) to the first one ($\hat{\lambda}_1^{ri} = 1$), by the following conversion formulas: $\hat{\gamma}_i^{ri} = \hat{\gamma}_i^{ve}\hat{\lambda}_1^{ve}$ and $\hat{\lambda}_i^{ri} = \hat{\lambda}_i^{ve}/\hat{\lambda}_1^{ve}$. Consequently, we can state that the choice between the two methods, at the net of convergence problems, does not affect the "qualitative" results of the analysis (i.e. the coefficients have different absolute values but their statistical significance, relative size, overall goodness of fit statistics, etc. are unaffected by the choice of the anchoring method). Moreover, we can always convert the estimated coefficients from one method to another, e.g. by fixing the coefficient of scale equals to the standard deviation of the variance of the structural equation calculated when we fix $\lambda_1^y = 1$ (i.e. $\hat{\lambda}_1^{ve} = (\hat{\psi}_1^{ri})^{0.5}$) and vice versa. As the other coefficients concern (i.e. intercepts and variances of measurement errors as well as the marginal effects of the observed "causes" on the "indicators", namely $\hat{\gamma}_i$ with $i = 6,7,8,9$ in Fig. 1), their values do not change regardless of the anchoring method applied.

[12] This is similar to Dybka et al.'s (2019) approach. They suggest to use the means and variances estimated in a Currency Demand Approach (CDA) to calibrate latent scores (estimated by CB-SEM) instead of anchoring the index on an arbitrary time period. However, Dybka et al.'s approach may be not appli-

Several studies deal with the issue of calibration of the MIMIC estimates and reviews of alternative methods have been recently published,[13] therefore, we focus on the two aspects that make the here proposed methods different from the current approaches.

First, while the existing methods aim to calibrate the latent scores of the MIMIC model, we propose two methods aimed to "estimate" the factor of scale in order to adjust estimated (structural) coefficients and, consequently, predict the latent scores. We also propose a third calibration approach providing the same results as in the first two methods that can be used as a robustness check to resolve inaccuracies in the previous methods (e.g. to discover "inverted trend" in predictions). By estimating the factor of proportionality of the SEM coefficients through these methods we present transparent algorithms to estimate both the IE and marginal effects of the drivers of the IE.

Second, with the exclusion of Dybka et al. (2019), the current calibration methods, can be used also with only one exogenous value of IE. We find that in order to achieve reliable estimates, at least two exogenous values of IE are required. Specifically, whatever indirect econometric method demands the additional assumption that IE is accurately estimated for "e" periods. According to our MC analysis, we find that the predictive performances of the MIMIC to estimate the IE when $e = 1$ is significantly worse than the predictions of IE based on $e \geq 2$.[14] As a consequence, we conclude that calibrating the MIMIC with 1 exogenous estimate leads to correct suitably predicted values only if each unit in the latent variable "IE" corresponds to a unit of change in the observable variable used as "reference indicator" (i.e. $\lambda_1^y = 1 = b_1^{tr}$). This may be a realistic hypothesis if as a "reference indicator" the scholar uses an accurate estimate of the IE. However, $\lambda_1^y = 1 = b_1^{tr}$ is as a special case of the more general approach covered by our calibration methods.

Accordingly, in the following sections, we assume to know 2 (and 3) consecutive[15] exogenous estimates of the IE. Moreover, to check the robustness of the predicted IE in terms of the chosen periods with available estimates of the IE, we estimate, for each MIMIC model, 15 different pairs (or 3 values) of external values, starting from the fairest periods ($t_0 - t_1$) in three sample sizes (25, 50 and 100 periods). In symbols, $\text{Inf}_{t^*}^{exog}$ indicates the exogenous estimates of the IE where $t^* \epsilon [w, w + e - 1]$ is the period used to calibrate the MIMIC output; $w = 1, \dots, 15$

Footnote 12 (continued)

cable if CDA is unfeasible (e.g. for the European countries adopting the euro, statistics on currency at country level are inaccurate since 2002 due to the changeover to new currency) or if the CDA is considered as an unreliable method to estimate the IE (e.g. if barter system or different currency are widespread).

[13] See e.g. Dell'Anno and Schneider (2006, 2009), Buehn and Schneider (2012), Breusch (2016), Dell'Anno and Davidescu (2019), Dybka et al. (2019), Medina and Schneider (2020).

[14] The statistical reason is that with only 1 exogenous value, we can estimate only 1 parameter. E.g. in the third method of Sect. 3.2.3 we could replace the "estimated" mean to the zero means of latent scores, while the variance of the estimate informality remains unknown.

[15] The method can be applied also when exogenous values are not consecutive. In this case simulations show that, on average, the goodness of fit improves as larger the chronological distance between exogenous values of the IE (D) is. To be cautious, we apply the worst situation, i.e. $D = 0$.

denotes the initial period used to define (rolling) "*window*" used to calibrate the latent scores in each simulated time series of the IE.[16]

The next section presents three calibration methods, that should be applied after the estimation of the SEM model by CB or PLS estimators. Taking into account that the source of bias of the MIMIC coefficients is the constraint on the coefficient of scale ($\lambda_1^y = 1$), the first two methods "adjust" the identification-bias of structural coefficients ($\hat{\gamma}$) to predict the IE in its real metric. In particular, they calculate the "true" values (a_i^{tr}) by rescaling the MIMIC coefficients by an exogenous (OLS) estimate of the reference indicator (b_1^{tr}) (i.e. $a_i^{tr} = \hat{\gamma}_i^{SEM} / \hat{\lambda}_1^{OLS}$ with $i =1,\dots,5$).[17] Once that all the structural coefficients are rescaled, the IE is predicted by multiplying them by the observed causes. The first two calibration methods differ on how to estimate b_1^{tr}. The third method produces the same estimates of IE as the first method, by calibrating directly the latent scores of the MIMIC.

2.2.1 First calibration method—adjusting structural coefficients by latent scores (structural model)

The first method estimates the measurement coefficient $\hat{\lambda}_1$ by minimizing the difference between latent scores estimated by the structural model and the exogenous values of the IE. The step-by-step procedure is:

Step 1 - Computing "first-stage" latent scores through structural coefficients - If we apply the CB estimator, the latent scores are obtained by:

$$\widehat{\text{Inf}}_t^{FS\_cb} = \sum_{i=1}^{5} {}^{cb}\hat{\gamma}_i x_{i,t}, \quad \text{with} \quad i = 1, \dots, 5; t = 1, \dots, T \tag{5a}$$

If we apply the PLS estimator, the latent scores are computed by:

$$\widehat{\text{Inf}}_t^{FS\_pls} = \sum_{i=1}^{5} {}^{pls}\hat{\gamma}_i \left(\sigma_{y_1} / \sigma_{x_i} \right) x_{i,t}, \quad \text{with} \quad i = 1, \dots, 5; t = 1, \dots, T \tag{5b}$$

Step 2 - Estimating the "coefficient of scale" by auxiliary OLS regressions - Taking into account that an approximation of Inf_t^{tr} is the exogenous estimate of the IE (Inf_{t*}^{exog}), then we estimate by OLS, the (inverse of) "true" value of the "coefficient of scale", i.e. $\hat{\lambda}_1 = E(b_1^{tr})$ and the intercept of the structural equation of the latent variable, $\hat{\gamma}_0 = E(a_0^{tr})$. Denoting with $= cb, pls$, the coefficients estimated by the CB and PLS algorithms are:

---

[16] E.g., with t=1,..,T, and $e=2$, the 1st iteration uses for calibration $Inf_{t=1}^{exog}$ and $Inf_{t=2}^{exog}$, the 2nd iteration $Inf_{t=2}^{exog}$ and $Inf_{t=3}^{exog}$, ..., the 15th and last iteration $Inf_{t=15}^{exog}$ and $Inf_{t=16}^{exog}$.

[17] It is also possible to correct identification-bias of measurement coefficients by applying: $\hat{b}_j^{tr} = \hat{\lambda}_j^{y-SEM} \hat{\lambda}_1^{OLS}$.

$$Inf_{t^*}^{exog} = \overbrace{^{ols\_est}\hat{\rho}_0}^{^{ols\_est}\hat{\gamma}_0} + \overbrace{^{ols\_est}\hat{\rho}_1}^{1/^{ols\_est}\hat{\lambda}_1} \overbrace{\widehat{Inf}_{t^*}^{FS\_est}}^{eq.5} + \varepsilon_{t^*} \quad with \; t^* \in [w, w+e-1] \tag{6}$$

where the estimated "coefficient of scale" is $^{ols\_est}\hat{\lambda}_1^* = 1/^{ols\_est}\hat{\rho}_1$.[18]

Step 3 - "Adjusting" SEM coefficients -

Sub-step 3.1: As the statistical relationships between "causes" and "IE" (a^{tr} in Eq. 1) concerns, we rescale the structural coefficients $^{est\_est}\hat{\gamma}_i^*$ by dividing the estimated coefficients ($^{est}\hat{\gamma}_i$) by $^{ols\_est}\hat{\lambda}_1^*$ estimated by step 2:

$$^{est}\hat{\gamma}_i^* = \frac{^{est}\hat{\gamma}_i}{^{ols\_est}\hat{\lambda}_1^*} \quad with \quad i = 1, \dots, 5; t = 1, \dots, T \tag{7}$$

Sub-step 3.2: As the statistical relationships between "IE" and "indicators" concern, (i.e. b^{tr} in Eqs. 2, 3 and 4), with the exclusion of the coefficient of scale that has been already estimated in step 2, the other measurement coefficits are "adjusted" by multiplying the original estimated values ($^{est}\hat{\lambda}_j$) by $^{ols\_est}\hat{\lambda}_1^*$:

$$^{est}\hat{\lambda}_j^* = {^{est}\hat{\lambda}_j}, {^{ols\_est}\hat{\lambda}_1^*} \quad with \quad j = 2, 3; t = 1, \dots, T \tag{8}$$

Sub-step 3.3: As the statistical relationships between "causes" and "indicators" concerns, (i.e. c^{tr} in Eq. 2, 3 and 4), since these coefficients are not affected by the anchoring ($\lambda_1^y = 1$), then $^{est}\hat{\gamma}_i^* = {^{est}\hat{\gamma}_i}$ with $i = 6,7,8,9$.

Step 4 - Calculate the "absolute" values of the Informal Economy -
The time series of the IE in "actual" metric ($^{est}\hat{IE}_t^{m\_1}$) is calculated by combining the results in steps 1, 2 and 3.1 as follows:

$$^{est}\hat{IE}_t^{m\_1} = \overbrace{^{ols\_est}\hat{\gamma}_0}^{step\,2} + \underbrace{\frac{\overbrace{\widehat{Inf}_t^{FS\_est}}^{step\,1}}{^{est}\hat{\gamma}_i^*}}_{step\,3.1} \quad with \quad i = 1, \dots, 5; t = 1, \dots, T \tag{9}$$

2.2.2 Second calibration method—adjusting structural coefficients by reference variable (measurement m.)

This second method estimates the measurement coefficient b_1^{tr} by regressing the exogenous estimate of the IE ($Inf_{t^*}^{exog}$) on the reference indicator (y_{1,t^*}) and calculates

---

[18] Equivalently, we can estimate the coefficient of scale ($^{ols\_est}\hat{\lambda}_1^*$) by $\widehat{Inf}_{t^*}^{FS\_est} = \rho_0 + {^{ols\_est}\hat{\lambda}_1^*}Inf_{t^*}^{exog} + \varepsilon_t$ and the intercept ($^{ols\_est}\hat{\gamma}_0$) as $^{ols\_est}\hat{\gamma}_0 = -\left({^{ols\_est}\hat{\rho}_0}/{^{ols\_est}\hat{\lambda}_1^*}\right)$.

the intercept of the structural model as the difference between means of latent scores and exogenous value of the IE. The step-by-step procedure is:

Step 1 - Estimating the "coefficient of scale" by auxiliary OLS regressions -
Estimating the "coefficient of scale" (i.e. $^{ols}\widehat{\lambda}_1^*$) through the first measurement equation (Eq. 2) after replacing the (unobserved) IE (Inf^{tr}) to the external estimate (i.e. $Inf_{t^*}^{exog}$)[19]:

$$y_{1,t^*} = \text{const} + {}^{ols}\widehat{\lambda}_1^* \text{Inf}_{t^*}^{exog} + \varepsilon_{t^*} \quad \text{with } t^* \in [w, w+e-1] \tag{10}$$

Step - 2 "Adjusting" SEM coefficients -
Rescaling the MIMIC coefficients $^{est}\widehat{\gamma}_i^*$ and $^{est}\widehat{\lambda}_j^*$, as in steps 3.1 and 3.2 of the first method by using:

$$^{est}\widehat{\gamma}_i^* = \frac{^{est}\widehat{\gamma}_i}{^{ols}\widehat{\lambda}_1^*} \quad \text{with } i = 1, \dots, 5; t = 1, \dots, T \tag{11}$$

$$^{est}\widehat{\lambda}_j^* = {}^{est}\widehat{\lambda}_j {}^{ols}\widehat{\lambda}_1^* \quad \text{with} \quad j = 2, 3; t = 1, \dots, T \tag{12}$$

Step 3 - Computing "first-stage" latent scores by adjusted structural coefficients
To compute "first-stage" latent scores ($\widehat{Inf}_t^{FS\_est}$) as in step 1 in the first method (Eq. 5a or 5b) using $^{est}\widehat{\gamma}_i^*$ computed by step 2 (Eq. 11).
Step 4 - Estimating the intercept of the structural model 1 ($^{\Delta\mu\_est}\widehat{\gamma}_0^*$) -
The intercept is calculated as the value that equalizes the means of the exogenous estimate of IE and the first-stage latent scores over the period t^*:

$$^{\Delta\mu\_est}\widehat{\gamma}_0^* = \text{Mean}\left(\text{Inf}_{t^*}^{exog}\right) - \text{Mean}\left(\widehat{\text{Inf}}_{t^*}^{FS\_est}\right) \quad \text{with} \quad t^* \in [w, w+e-1] \tag{13}$$

Step 5 - Calculating the "absolute" values of the Informal Economy -
Predicting the IE by using latent scores of step 3 and the intercept of step 4:

$$^{est}\widehat{IE}_t^{m\_2} = \overbrace{^{\Delta\mu\_est}\widehat{\gamma}_0^*}^{eq.13} + \widehat{\text{Inf}}_t^{FS\_est}, \quad \text{with} \quad t = 1, \dots, T \tag{14}$$

---

[19] The regression 10 does not include the effect of x_6 on y_1 as in Eq. 2. This omission of a relevant variable is due to the hypothesis to use the lowest possible number of exogenous estimates of the IE (e=2). In general, if the omitted variable is not correlated with the included regressor ($Inf_{t^*}^{exog}$), the estimate of $E(b_1^{tr}) = {}^{ols}\widehat{\lambda}_1$ by Eq. 12 remains unbiased. On the contrary, if $Corr(Inf, x_6) \neq 0$ then we need to include the omitted variable to get an unbiased estimate of $^{ols}\widehat{\lambda}_1$. Given that this possibility, depends on data availability (e>3), we simulate the effect of this inclusion later in the MC analysis.

2.2.3 Third calibration method—adjusting latent scores by estimating the mean and standard deviation of the IE

This method deals with the issue of indeterminacy by directly constraining the mean and the standard deviation of the latent scores.[20] This method generates the same predicted values as in the first method therefore we omit reporting these findings in the MC analysis.[21] The third calibration method consists of four steps:

Step 1 - Computing "first-stage" latent scores by the structural coefficients - (See step 1 of the first method, section 2.2.1)

Step 2 - Computing "second-stage" latent scores -
Standardizing the latent scores ($\widehat{Inf}_t^{FS\_est}$) to get a new (second-stage) time series of the latent scores with zero mean and unit standard deviation ($\widehat{Inf}_t^{z-FS\_est}$):

$$\widehat{Inf}_t^{z-FS\_est} = \left(\widehat{Inf}_t^{FS\_est} - \widehat{\mu}_{\widehat{Inf}_t^{FS\_est}} \right) / \widehat{\sigma}_{\widehat{Inf}_t^{FS\_est}}, \quad \text{with} \quad t = 1, \dots, T \tag{15}$$

Step 3 - Minimizing the difference between predicted IE and exogenous values -
Estimating the mean ($^{ols\_est}\widehat{\mu}_{IE}$) and standard deviation ($^{ols\_est}\widehat{\sigma}_{IE}$) which minimize (by OLS) the difference between "second-stage" (standardized) latent scores and exogenous estimates of the IE (Inf_{t*}^{exog})[22]:

$$Inf_{t*}^{exog} = \overbrace{^{ols\_est}\widehat{\phi}_0}^{^{ols\_est}\widehat{\mu}_{IE}} + \overbrace{^{ols\_est}\widehat{\phi}_1}^{^{ols\_est}\widehat{\sigma}_{IE}} \widehat{Inf}_{t*}^{z-FS\_est} + \varepsilon_{t*}, \quad \text{with} \quad t* \in [w, w+e-1] \tag{16}$$

Step 4 - Calculating the "absolute" values of the Informal Economy -
The IE in "actual" metric ($^{est}\widehat{IE}_t^{m\_3}$) is calculated by un-standardizing, the (standardized) "second-stage" latent scores over the full period ($t = 1, \dots, T$):

---

[20] A simplified version of this method may be applied if only 1 external value of the IE (i.e. $e = 1$) is available. It consists of adding a constant to the latent scores in order to constraint the "absolute" values of the IE equal to the external estimate at time t* (i.e. $^{CB}\widehat{IE}_t^* = \varpi + \widehat{Inf}_t^{est}$ with $\varpi = Inf_{t*}^{exog} - \widehat{Inf}_{t*}^{est}$) as in Eq. 13. However, this calibration has poor predictive performances and cannot be applied to standardized latent scores (i.e. it is inapplicable to PLS-SEM). Accordingly, we report simulations only focusing on $e = 2, 3$.

[21] In general, latent scores can be estimated by measurement or structural models. In the MIMIC specification they should be theoretically equal, because they refer to the same construct. However in empirical applications some minor differences can emerge. In this analysis we find that in CB-SEM these differences are negligible, in PLS the decision to use the outer model instead of the inner model to predict latent scores has more consequence on the scores. For the sake of comparability among calibration methods, we use latent scores predicted by the inner model in CB-SEM and PLS-SEM. The third calibration can be easily applied to both types of latent scores by replacing the "first-stage" latent scores. Another difference between this method and the two previous calibration approaches is that it doesn't adjust the estimated coefficients therefore, in order to estimate the marginal effects of the "causes" on the size of the IE, scholars may estimate Eq. 1 by replacing the dependent variable with the adjusted latent scores (i.e. $Inf^{tr} = {}^{est}\widehat{IE}_t^{m\_3}$).

[22] This specification derives to inverting the formula of standardization, i.e. from $z_t = (x_t - \widehat{\mu})/\widehat{\sigma}$ to $x_t = \widehat{\mu} + \widehat{\sigma} z_t$.

$$est\,\widehat{IE}_t^{m\_3} = \overbrace{^{ols\_est}\widehat{\mu}_{IE}}^{eq.16} + \overbrace{^{ols\_est}\widehat{\sigma}_{IE}}^{eq.16}\overbrace{\widehat{\mathrm{Inf}}_t^{z-FS}}^{eq.15}, \quad \text{with} \quad t = 1, \dots, T \tag{17}$$

3 MC analysis of predictive performances of the MIMIC approach

This section compares the reliability of the MIMIC approach and the performances of the calibration methods under two strongly correlated perspectives: the reliability of estimated coefficients and the goodness-of-fit of predicted IE. Moreover, in order to have a benchmark of SEM capacity to estimate "true" coefficients, we also compare MIMIC outputs to OLS estimates of regressions 1–4. The MC analysis includes different scenarios[23]:

Two model specifications: the "classical" MIMIC specification, labeled as *Model A*, and a more complex *Model B*. The "true" coefficients are fixed to obtain simulated values of the IE, indicators and causes that range over realistic domains.[24]

Two probability distributions of observed variables. The pseudo-random numbers used for observable causes (x) are distributed according to a normal or uniform probability distribution. This robustness check is included for three reasons: (a) if the variables are not (multivariately) normally distributed, then it is possible for ML estimators (i.e. CB-SEM), to produce biased standard errors and an ill-behaved "chi-square" test of the overall model fit[25]; (b) one of the main comparative advantages of the PLS estimator over CB-SEM consists in better performances in presence of nonnormal data; (c) The variables usually employed in this type of MIMIC application are non-normal distributed, therefore removing this assumption increases the external validity of the MC analysis.

Two sets of exogenous values of the IE. All calibration methods require e exogenous values of the "actual" IE. We assume the availability of two and three values for calibration ($e = 2, 3$).[26]

---

[23] Online Appendix B shows: a Tree diagram of the structure of the MC simulation (Fig. B.1); the "true" values of coefficients and the descriptive statistics of the simulated variables (Table B.1).

[24] As the CB-SEM often encounters convergence problem of algorithm that minimizes the ML fitting function, to set the "true" coefficients, we exclude those values that cause considerable convergence problem for CB estimator. The problem of non-convergence is a relevant issue for the MIMIC approach and can be faced in different ways. In this research, we reduce the non-convergence rates by switching, every 10 iterations, four different algorithms to maximize log-likelihood fitting function (namely, Broyden–Fletcher–Goldfarb–Shanno, Berndt–Hall–Hall–Hausman, Davidon–Fletcher–Powel and Newton–Raphson). Once that the CB-SEM converges, we use the estimated model-implied covariance matrix as the starting values for the Newton–Raphson algorithm applied to get final estimation of the CB-MIMIC. See Gould et al. (2010) for details on the algorithms used to ML estimation.

[25] However, as we show in Sect. 4 some corrections are available if there is a violation of multivariate normality assumption, e.g. Satorra-Bentler estimator (see for detail Bollen, 1989).

[26] Intuitively, as e increases, the reliability of calibration methods improves, up to $e = T$, where the estimation of the IE by MIMIC has not practical significance. A further benefit to use more than 2 exogenous estimates of the IE, is that we can estimate confidence intervals for each calibration method and consequently, estimate a range of IE predictions. This can be a further extension of this research, however as hints for these more in-depth studies, we see that with $e = 3$ the confidence intervals are still too wide to provide practical indications on the "actual" size of the predicted IE.

Three sample sizes. We consider three sample size dimensions (i.e., $T = 25$, 50, 100) to check the properties of estimators (consistency and convergence rates). According to this structure, there are 1200 simulated pseudo-random datasets (i.e. 200 based on normal and 200 on uniformly distributed pseudo-random "causes" both with three different dimensions of time series, $T = 25$, 50, 100). Each of the 1200 datasets, is estimated by a MIMIC model calibrated by a rolling window method (i.e. we consider 15 different consecutive sets of exogenous values of IE) and replicate the analysis in two scenarios i.e. with $e = 2$ and 3. It implies estimating 36,000 models "A" and 36,000 models "B". Each model A is estimated by OLS (assuming that IE is observable), CB-SEM and PLS-SEM estimators. As the specification of the MIMIC B concerns, given that it cannot be estimated by the PLS-SEM, we report CB-SEM (calibrated) coefficients and predictions of the IE.

3.1 How reliable are the MIMIC estimates?

In the first step of the MC analysis, we compare the estimated coefficients of regressions (1–4) obtained by OLS, CB-SEM and PLS-SEM. In a real empirical application, it is not possible to estimate regressions 1–4 by OLS because IE is unobservable. In this analysis, we use the simulated values of Inf^{tr} as calculated by Eq. 1. Estimators are evaluated in terms of (un)biasedness,[27] efficiency[28] and consistency.[29]

In order to disconnect the assessment of SEM estimators from the reliability of the calibration methods, we initially "adjust" the estimated coefficients by the "true" value of the coefficient of scale (i.e. $\hat{a}_i = \hat{\gamma}_i^{est}/b_1^{tr}; \hat{b}_j = \hat{\lambda}_j^y b_1^{tr}$, where $b_1^{tr} = 5$). Table 1 reports the mean of the indexes of bias, efficiency and consistency separately for each model, based on the smallest sample (T = 25).

As expected,[30] OLS dominates SEM estimators, but no relevant differences emerge among estimators. In relative terms, the CB has better performances than the PLS in terms of mean-unbiasedness, efficiency, and consistency. This result is consistent with the predominant literature that argues as, if parametric statistical assumptions are satisfied, the CB-SEM is superior to the non-parametric approach

---

[27] An estimator is unbiased if the mean of the sampling distribution of the estimator is equal to the true parameter value. We assess this property by the root mean squared error (*rmse*): $rmse\left(\hat{\theta}\right) = \left[\frac{1}{N} \sum_{i=1}^{N} \left(\hat{\theta}_i - \theta^{tr}\right)^2\right]^{0.5}$.

[28] An estimator is efficient if there is a small deviance between the estimated value and the "true" value. We evaluate this property by comparing the variance per estimated coefficient, i.e. $Eff\left(\hat{\theta}\right) = 1000 * \left(\sigma^2_{\hat{\theta}_i}/N\right)$, where N indicates the number of estimated coefficients. We divide the variance of estimated coefficients by the number of estimated coefficients, because N changes between CB and PLS (or OLS) estimators due to non-convergence of ML estimator.

[29] An estimator is consistent if, as the sample size increases, the sampling distribution of the estimator becomes increasingly concentrated at the true parameter value. We measure it by the difference in bias, measured by *rmse*, as sample size increases by 100%: $Cons^{\Delta T_{25-50}} = rmse\left(\hat{\theta}\right)^{T=50} - rmse\left(\hat{\theta}\right)^{T=25}$ and $Cons^{\Delta T_{50-100}} = rmse\left(\hat{\theta}\right)^{T=100} - rmse\left(\hat{\theta}\right)^{T=50}$.

[30] Because of the "true" values of the latent variable (Inf^{tr}) are included as dependent variable in OLS regressions.

(i.e., PLS). However, the advantages of the PLS over the CB algorithm should be assessed taking into account the substantially better performance of the PLS in terms of convergence rate (see the last column of Table 1). While the PLS estimator always converges, the CB algorithm does not converge in about 40% of the models.[31]

Once the reliability of the MIMIC model has been verified when an exogenous estimate of the "coefficient of scale" is provided, we assess if calibration methods provide a reliable estimate of b_1^{tr}.

Table 2 reports results on the most demanding scenario in calibration methods, i.e. a small sample size ($T=25$) and the lowest numbers of exogenous values of the IE, i.e. $e=2$ (see Online Appendix C for results based on $e=3$). By focusing on the structural coefficients, we infer that: (a) the estimates of the "classical" MIMIC specification are more accurate than the estimates based on the (unconstrained)[32] model B. This result is consistent with Dell'Anno and Schneider's (2006) statement, of a trade-off between the "statistical" goodness-of-fit provided by parsimonious MIMIC specifications and the exhaustiveness of the economic theory behind the MIMIC model. (b) As the choice of the best calibration method concerns, it depends on the model specification. Specifically, the first calibration method has better performance in estimating more complex measurement models while the second method is preferable when estimating the "classical" model A. (c) As the choice of SEM estimator concerns, in general, CB overperforms PLS. However, some differences emerge in terms of efficiency as a function of calibration methods. When applying the first method, CB dominates PLS in terms of mean squared error and efficiency. On the contrary, when the second calibration method is applied, the CB dominates in terms of mean-unbiasedness but has lower efficiency than PLS. (d) As the number of exogenous values of the IE concerns, the estimates based on $e=2$ (Table 2) are worse, in terms of mean-unbiasedness and efficiency than estimates based on $e=3$ (see Online Appendix C). Moreover, while with $e=2$ both the SEM estimators are often not consistent, with three exogenous values this lack of consistency disappears; (e) In addition, with $e=3$, although the ranking of calibration methods and estimators is preserved in terms of statistical properties, the (absolute) differences in performances among calibration methods and estimators decrease.

In conclusion, we find that on average, the "calibrated" estimates of MIMIC coefficients are reliable because they provide unbiased estimates of the "true" coefficients and sufficient levels of efficiency also in small sample ($T=25$). The choice of a calibration strategy depends on the model's complexity. The second method is preferable for the "classical" MIMIC model, while for the more "complex" measurement model of the "reference" indicator, the first method has better performances. As the choice of

---

[31] Table B.2 in Online Appendix B shows the ML convergence rates. We find that convergence increases as the sample increases; as model specification becomes more parsimonious (i.e. model A) and it is not affected by the probability distribution of the observed variables (i.e. normal versus uniform distribution).

[32] The model with more restrictions or less free parameters (i.e., more degrees of freedom)—also defined as "reduced" model—is nested within the less restricted model, which is defined as "full" model. Accordingly, Model A is nested in Model B because the estimated (free) parameters in Model A are a subset of the free parameters in Model B.

Table 1 Coefficients Scaled by the "TRUE" coefficient of scale ($b_1^{tr} = 5$)

Coeff	True	Mode	Mean			Mean-bias (rmse)			Efficency			Consistency			Conv.
			OLS	CB	PLS	OLS	CB	PLS	OLS	CB	PLS	OLS	CB	PLS	CB
a_1	-1	A	-1.012	-1.008	-1.024	0.155	0.164	0.283	0.009	0.018	0.028	-0.049	-0.056	-0.097	61.2%
		B	-1.012	-1.012		0.155	0.167		0.009	0.021		-0.049	-0.061		52.3%
a_2	2	A	1.997	2.000	2.029	0.161	0.169	0.457	0.011	0.020	0.072	-0.054	-0.061	-0.140	61.2%
		B	1.997	2.008		0.161	0.183		0.011	0.027		-0.054	-0.069		52.3%
a_3	2.5	A	2.504	2.509	2.532	0.061	0.062	0.262	0.001	0.002	0.023	-0.021	-0.021	-0.060	61.2%
		B	2.504	2.510		0.061	0.062		0.001	0.003		-0.021	-0.021		52.3%
a_4	1	A	0.998	0.992	1.007	0.081	0.087	0.216	0.003	0.005	0.016	-0.027	-0.033	-0.061	61.2%
		B	0.998	0.993		0.081	0.089		0.003	0.006		-0.027	-0.033		52.3%
a_5	-1	A	-0.987	-0.996	-1.013	0.169	0.175	0.284	0.011	0.021	0.029	-0.063	-0.063	-0.099	61.2%
		B	-0.987	-1.005		0.169	0.172		0.011	0.023		-0.063	-0.066		52.3%
b_1	5	A	5.001	5.000	5.000	0.021	0.000	0.000	0.000	0.000	0.000	-0.008	0.000	0.000	61.2%
		B	5.001	5.000		0.022	0.000		0.000	0.000		-0.008	0.000		52.3%
b_2	1	A	1.000	1.000	1.001	0.020	0.021	0.009	0.000	0.000	0.000	-0.006	-0.006	0.000	61.2%
		B	1.000	1.000		0.020	0.022		0.000	0.000		-0.006	-0.008		52.3%
b_3	.7	A	0.700	0.702	0.703	0.019	0.019	0.010	0.000	0.000	0.000	-0.007	-0.007	0.000	61.2%
		B	0.700	0.703		0.021	0.021		0.000	0.000		-0.008	-0.009		52.3%
c_1	1	B	0.992	1.004		0.153	0.157		0.009	0.018		-0.057	-0.062		52.3%
c_2	-1	B	-1.001	-0.996		0.081	0.082		0.003	0.005		-0.028	-0.026		52.3%
c_3	-.7	B	-0.703	-0.684		0.166	0.160		0.011	0.020		-0.066	-0.059		52.3%
c_4	-2	B	-2.004	-1.986		0.034	0.124		0.001	0.014		-0.011	-0.045		52.3%

Table 2 Assessment of Calibrated Coefficients ($T = 25$ and $e = 2$)

	M.	Mean					Bias (Mean)					Efficiency					Consistency				
		ols	cb-1	cb-2	pls-1	pls-2	ols	cb-1	cb-2	pls-1	pls-2	ols	cb-1	cb-2	pls-1	pls-2	ols	cb-1	cb-2	pls-1	pls-2
a_1	A	-1.012	-0.993	-1.014	-0.849	-1.015	0.155	4.540	1.107	7.518	1.24	0.009	11.46	0.712	27.43	0.565	-0.049	2.390	0.394	.374	-0.13
	B	-1.012	-0.835	-1.088			0.155	9.585	18.78			0.009	84.53	372.9			-0.049	-4.099	4.473		
a_2	A	1.997	1.990	2.019	1.724	2.028	0.161	10.87	2.41	14.12	2.834	0.011	68.06	3.598	93.80	3.287	-0.054	6.299	0.509	-.005	-0.56
	B	1.997	1.652	2.121			0.161	18.49	35.46			0.011	299.4	1345			-0.054	-4.056	15.7		
a_3	A	2.504	2.494	2.528	2.136	2.518	0.061	12.94	3.094	18.3	2.683	0.001	95.79	5.873	155.0	2.598	-0.021	7.709	0.392	.717	.219
	B	2.504	2.056	2.728			0.061	23.18	43.58			0.001	478.6	2012			-0.021	-5.025	16.26		
a_4	A	0.998	0.993	0.997	0.855	0.998	0.081	5.643	1.139	6.434	1.143	0.003	18.51	0.770	19.47	0.494	-0.027	2.850	0.315	.400	.038
	B	0.998	0.814	1.038			0.081	9.141	18.03			0.003	72.43	344.1			-0.027	-2.942	5.03		
a_5	A	-0.987	-0.991	-1.000	-0.877	-1.007	0.169	5.455	1.271	6.457	1.315	0.011	16.8	0.973	19.44	0.649	-0.063	3.502	0.112	.973	-.042
	B	-0.987	-0.862	-1.099			0.169	8.302	17.03			0.011	59.63	313.2			-0.063	-1.353	7.585		
$\bar{a}$	A	0.700	0.699	0.706	0.598	0.704	0.125	7.891	1.804	10.57	1.843	0.007	42.12	2.385	63.03	1.519	-0.043	4.550	0.344	.492	-.095
	B	0.700	0.565	0.740			0.125	13.74	26.58			0.007	198.9	877.5			-0.043	-3.495	9.808		
b_1	A	5.001	5.69	4.842	5.392	4.918	0.021	18.49	5.49	23.04	4.553	0.000	254.7	25.79	215.5	10.04	-0.008	26.69	4.93	70.2	10.16
	B	5.001	5.328	6.061			0.022	18.24	63.5			0.000	280.3	2914			-0.008	26.43	349.8		
b_2	A	1.000	1.14	0.969	1.079	0.984	0.020	3.751	1.113	4.601	0.912	0.000	10.52	1.064	8.60	0.403	-0.006	5.317	0.973	14.1	2.024
	B	1.000	1.068	1.216			0.020	3.715	12.99			0.000	11.67	122.8			-0.006	5.247	69.55		
b_3	A	0.700	0.799	0.679	0.759	0.692	0.019	2.608	0.788	3.274	0.648	0.000	5.05	0.533	4.38	0.204	-0.007	3.774	0.676	9.91	1.426
	B	0.700	0.752	0.861			0.021	2.668	9.303			0.000	6.031	63.1			-0.008	3.637	48.65		
$\bar{b}$	A	2.234	2.543	2.163	2.410	2.198	0.020	8.282	2.464	10.30	2.038	0.000	90.09	9.129	76.14	3.548	-0.01	11.93	2.193	31.4	4.536
	B	2.234	2.383	2.713			0.021	8.209	28.6			0.000	99.33	1033			-0.01	11.77	156		
c_1	B	0.992	1.004	1.004			.153	.157	.157			.009	.018	.018			-.057	-.062	-.062		
c_2	B	-1.001	-0.996	-0.996			.081	.082	.082			.003	.005	.005			-.028	-.026	-.026		
c_3	B	-0.703	-0.684	-0.684			.166	.160	.160			.011	.02	.02			-.066	-.059	-.059		
c_4	B	-2.004	-1.986	-1.986			.034	.124	.124			.001	.014	.014			-.011	-.045	-.045		
$\bar{c}$	A	-0.679	-0.666	-0.666			.109	.131	.131			.006	.014	.014			-.041	-.048	-.048		

Notes: $\bar{a}$, $\bar{b}$, and $\bar{c}$ denotes the means. OLS estimates and the estimates of "c" coefficients do not need calibration, therefore, are equal to those reported in Appendix C.

SEM estimators concerns, CB is the first-best solution, but taking into account that the ML estimator often does not converge, then PLS is a worthwhile alternative.

3.2 How reliable are the MIMIC predictions of the informal economy?

In this section, we assess the goodness-of-fit by using the simulated values of the IE as the benchmark for the predicted values. The goodness-of-fit is analyzed under four complementary perspectives.

3.2.1 Descriptive statistics on the distribution of the IE

Table 3 reports the averages of the mean, standard deviation, median, minimum and maximum values of predicted time series of the IE assuming calibration with $e = 2$.[33] This analysis shows that while both the calibration methods provide accurate predictions of mean and median of the true IE, some relevant differences between the two methods emerge in terms of variance of IE. In particular, by calibrating the MIMIC by the structural model (first method), we observe an overestimation of the "true" standard deviation, especially for uniform distribution. As a consequence of these (upward) biased estimates of variance, there are outliers in the predicted series (e.g. minimum and maximum values of the predicted values out of the normal range of informality ratio 0–100). In line with the findings of the previous sections, the lower accuracy to estimate coefficients in the more complex MIMIC specification leads to lower accuracy in the prediction of the IE based on Model B than model A.

3.2.2 Analysis of the distributions of variance explained by predicted IE and presence of outliers

We report the most indicative statistics of the distribution of the fraction of variance explained[34] by the MIMIC predictions, in different scenarios and based on the smallest sample size. In particular, Table 4 shows the statistical reliability of the predicted IE by highlighting the frequency of models with $R^2 > 0.95$.

According to MC findings, $R^2 > 0.95$ in more than 95% of the predicted IE based on the MIMIC model A when it is calibrated by the second calibration approach by using $e = 2$. Moreover, increasing the exogenous values for calibration ($e = 3$) the goodness of fit furtherly increases ($R^2 > 0.99$) and achieves a noteworthy level ($R^2 > 0.95$) also by applying the first calibration method. In detail, the best estimator for MIMIC A is the CB estimator calibrated by the second method and, if the CB algorithm does not converge (it happens for 40–50% of the models), the PLS estimator calibrated by the second method is the best alternative.

The first calibration method produces the best predictions of the IE for MIMIC B. However, for this specification, the coefficient of determination is greater than 0.95 only in 90% of the estimated models. As the effect of a larger number of exogenous values ($e = 3$) concern, we observe at least two beneficial effects: an improvement of the overall goodness-of-fit and a decrease in the differences between SEM estimators and calibration methods.

---

[33] We don't report descriptive statistics based on e = 3 for the sake of brevity. However they are more accurate than $e = 2$.

[34] R^2 is calculated as 1-the ratio of the residual sum of squares (RSS) and the total sum of squares (TSS). In symbols, $R^2 = 1 - \left[\sum_{i=1}^{N} \left(Inf_t^{tr} - {}^{est}\hat{IE}_t^{m\_c} \right)^2 / \sum_{i=1}^{N} \left({}^{est}\hat{IE}_t^{m\_c} - \overline{Inf}_t^{tr} \right)^2 \right]$ with $est = OLS, CB, PLS$; and $m\_c = m\_1, m\_2$. Results based on larger sample sizes are provided in Appendix D.

The MC analysis points out that the calibration procedures may generate some significant inaccuracies in the predictions. In order to exclude these (few)[35] cases, a simple rule of thumb, which we label as "Min–Max" in Table 4, is applied. It consists of excluding those models that generate at least one negative or larger than 100% predictions of the IE, i.e. $^{est}\widehat{IE}_t^{m\_c} \leq 0$ or $^{est}\widehat{IE}_t^{m\_c} \geq 100$.[36] We find that this "Min–Max" rule of thumb is sufficient to drastically decrease standard deviations of estimated R^2 as well as improve the goodness of fit, therefore may effectively reduce the risk to have confidence in inaccurate MIMIC predictions in real-world applications.

3.2.3 Graphical analysis of predicted IE and the issue of inverted trend

The third analysis of the reliability of the MIMIC estimates is based on graphical comparisons of predictions in different scenarios (see Online Appendix E, Figs. E.1, for $e = 2$ and Figs. E.2 for $e = 3$).[37]

The graphical analysis corroborates the previous conclusions[38] and allows to identify some cases where the impreciseness of predicted values is relevant.[39] Following Andrews et al. (2020) we run a *reverse sensitivity analysis* to show in which cases the SEM approach reverses the predicted trend of the IE. Precisely, we find that calibration methods could lead to both a biased estimate of the standard deviation of the predicted IE and a wrong sign of the factor of scale, which causes an inverted trend of the predicted values with respect to the "true" values.[40] In both cases, some solutions are possible.

As the first issue concerns, given that significant bias in estimated standard deviation generates outliers (i.e. values out of range 0–100 of the informality ratio) in predicted values, by applying the "Min–Max" criterion proposed above we are able to identify these wrong predictions.

---

[35] For instance, column "N" of Table 4 shows as—with the exclusion of Model B calibrated by the second method where the predicted IE is out of the range 0–100% in 9.6% of the estimated models—the presence of a (potential) outlier in the predicted IE occurs in less than 1% of predicted IE.

[36] This criterion to detect outliers is consistent with the common practice to define the latent variable of the MIMIC model as a ratio (e.g. IE on GDP, value-added, labor force, etc.).

[37] We select four simulated series of IE whose the CB estimator converges over all 24 combinations of sample size, probability distribution, model specifications and calibration methods, (i.e., 3rd, 45th, 48th and 78th).

[38] Specifically: the MIMIC model usually produces reliable predictions; the second calibration method is frequently the best method to calibrate model A; the first calibration method outperforms the second one in predicting model B; goodness-of-fit strongly improves with 3 exogenous values of the IE; CB and PLS have similar performances with CB which is, often but not always (e.g. Fig. E.1c simulations #3 and #48 calibrated with t = 1, 2), better than PLS.

[39] See Model A, simulation #3, estimated by PLS, calibrated by the first method and over the period t* = 1, 2 (Fig. E.1.a) or Model B, simulation #3 estimated by CB_2 and calibrated with t* = 5, 6 (Fig. E.1.c).

[40] See Model A and B, simulation #45, calibrated by the first method and over the period t* = 1, 2 (Figs. E.1.b and E.1.d) or simulation #3, estimated by CB, calibrated by the second method and over the period t* = 15, 16 (Fig. E.1.d).

As the inverted trend concerns, although the best solution is to increase the accuracy of calibration by increasing the number of exogenous values,[41] it may be inapplicable due to data unavailability. An alternative solution is to check if the sign of "coefficient of scale" ($\widehat{\lambda}_1^*$) as estimated by the calibration equation is incorrect. This check may be done in different ways. In general, the graphical analysis may provide sufficient evidence in recognizing the inverted trend (see e.g. Fig. E.1.b,d in Online Appendix E). However, in real-world applications with few exogenous values of the IE, this approach may be inconclusive.

A second strategy is to apply the third calibration approach. This method generates the same quantitative results as the first one, but by inverting the formula of standardization (see footnote 22). This means that by checking the sign of the coefficient $^{ols\_est}\widehat{\rho}_1$ in regression 16, a scholar may uncover if the trend is reversed.[42]

A third check for the presence of an inverted trend may be based on Dell'Anno's (2003) suggestion. He suggested exploiting the theoretical and empirical knowledge on the relationship between IE and observed variables (indicators and causes). For instance, if the signs of the "adjusted" estimated structural and measurement coefficients predominantly diverge from "unquestionable" theories and/or empirical evidence, then this knowledge, combined with the previous checks, validate the hypothesis of a reversed sign the scale factor (i.e. $-1 * {}^{ols}\widehat{\lambda}_1^*$).[43] We explore the incidence of this issue and report results in Online Appendix F.[44] In short, although an inverted trend in MIMIC predictions could emerge, it may be recognized by combining graphical, statistical and theoretical checks. Once that inverted trend is ascertained,

---

[41] In Online Appendix E, Fig. E.2 points out as, with e = 3, the inverted trends are fixed (e.g. compare Figs. E.1.b and E.1.d with E.2.b and E.2.d). However, by increasing the number of exogenous estimates of the IE does not guarantee the solution to this issue. For instance, in our MC simulation, also with e = 3, there are some (few) cases (see Table F.1) where the sign of the estimated $^{est}\widehat{\lambda}_1$ is still wrong (i.e. negative).

[42] In other words, given that $^{ols\_est}\widehat{\rho}_1$ is an estimate of the standard deviation of the IE, if it has a negative sign, the trend of predictions will be inverted respect to the "true" values. If it is the case, the researcher may change the sign of the "coefficient of scale" ($^{ols\_est}\widehat{\rho}_1$ in Eqs. 6 and 16) to calculate the right trend of the IE.

[43] For instance, if the reference indicator is an exogenous estimate of the IE, then an "unquestionable" empirical evidence is that predicted IE and exogenous estimates of the IE are positively correlated. This implies that if the "coefficient of scale" estimated by the first measurement equation (i.e. Equation 10) is negative, the scholar has a clear signal that the second calibration method generate predicted IE with an inverted trend.

[44] Considering the most demanding case (i.e. minimum sample size t = 25 and only two exogenous estimates of IE e = 2): (a) the probabilities of a wrong sign in $\widehat{\lambda}_1^*$ is sufficiently low. Specifically, for Model A, it ranges between 0.6% estimated by PLS (second calibration approach) to 3.5% estimated by PLS (first calibration method), for model B (estimated by CB) it ranges between 2.7% if we apply the first calibration method to 7.8% when the second calibration method is allied (See Fig. F.1 for full details of the simulations based on $t = 25, 50, 100$ and $e = 2$ and 3). (b) As the usefulness to predict the IE by applying both SEM estimators and different calibration methods concerns, we find that the combined probability that the predicted IE have an inverted trend in a model estimated by the same SEM estimator but with 2 calibration methods (i.e. $CB^{1st} \cap CB^{2nd} = 9/3675 = 0.24\%$) or by applying the same calibration method with both SEM estimators (i.e. $CB^{1st} \cap PLS^{1st} = 60/3675 = 1.63\%$) are negligible. See Table F.2 for other possible combinations of SEM estimators, calibration methods, MIMIC specifications, sample sizes and $e = 2$ or 3.

Table 3 Descriptive Statistics of Predicted IE (Mean, St. Dev., Min, Median, Max) and $e=2$

	Specification →	Model A						Model B					
	Distr. of x →	Normal			Uniform			Normal			Uniform		
	Sampe Size →	25	50	100	25	50	100	25	50	100	25	50	100
Mean	*True*	*36.58*	*36.62*	*36.54*	*52.12*	*52.27*	*52.35*	*34.58*	*34.62*	*34.54*	*52.12*	*52.27*	*52.35*
	Ols	36.58	36.62	36.54	52.12	52.27	52.35	34.58	34.62	34.54	52.12	52.27	52.35
	cb, 1st meth.	36.11	38.93	36.86	54.41	51.62	52.13	33.88	32.80	33.64	65.17	51.81	52.48
	cb, 2nd meth.	36.24	36.65	36.26	52.57	52.74	52.65	34.38	32.70	33.96	63.16	61.00	61.85
	pls, 1st meth.	38.11	36.17	36.47	52.07	53.56	52.83						
	pls, 2nd meth.	36.52	36.70	36.45	52.12	52.57	52.45						
Standard Dev.	*True*	*8.12*	*8.10*	*8.11*	*16.96*	*17.10*	*17.05*	*8.12*	*8.10*	*8.11*	*16.96*	*17.10*	*17.05*
	Ols	8.07	8.05	8.05	16.93	17.07	17.02	8.07	8.05	8.05	16.93	17.07	17.02
	cb, 1st meth.	11.09	12.85	10.64	20.09	18.03	19.13	10.73	12.26	15.04	25.51	18.17	19.09
	cb, 2nd meth.	8.53	8.35	10.17	17.44	17.96	17.12	10.66	17.28	12.63	56.44	53.45	49.41
	pls, 1st meth.	13.76	11.05	12.16	20.15	22.77	18.39						
	pls, 2nd meth.	8.73	8.44	9.66	17.72	17.82	17.09						
Min	*True*	*20.59*	*18.26*	*16.00*	*20.76*	*17.99*	*15.38*	*18.59*	*16.26*	*14.00*	*20.76*	*17.99*	*15.38*
	Ols	20.54	18.42	16.10	20.81	17.99	15.55	18.54	16.42	14.10	20.81	17.99	15.55
	cb, 1st meth.	14.19	10.27	9.75	16.60	15.41	10.90	12.56	4.48	-5.45	17.71	15.17	11.40
	cb, 2nd meth.	19.40	17.54	9.10	20.17	16.79	15.53	12.75	-6.36	1.55	-46.70	-43.38	-50.59
	pls, 1st meth.	9.42	9.95	5.25	14.78	7.63	13.02						
	pls, 2nd meth.	19.16	17.59	11.18	19.37	16.90	15.48						
Median	*True*	*36.42*	*36.59*	*36.58*	*51.89*	*52.24*	*52.23*	*34.42*	*34.59*	*34.58*	*51.89*	*52.24*	*52.23*
	Ols	36.39	36.58	36.54	51.82	52.24	52.30	34.39	34.58	34.54	51.82	52.24	52.30
	cb, 1st meth.	35.72	38.65	36.85	54.31	51.52	52.17	33.71	32.47	34.17	65.18	51.73	52.31
	cb, 2nd meth.	35.96	36.55	36.30	52.24	52.62	52.60	34.21	33.04	33.97	65.38	60.52	60.81
	pls, 1st meth.	37.90	36.15	36.57	51.77	53.05	52.76						
	pls, 2nd meth.	36.30	36.68	36.47	51.79	52.51	52.38						
Max	*True*	*52.84*	*55.03*	*56.99*	*83.21*	*87.05*	*89.27*	*50.84*	*53.03*	*54.99*	*83.21*	*87.05*	*89.27*
	Ols	52.69	54.95	56.80	83.24	86.98	89.20	50.69	52.95	54.80	83.24	86.98	89.20
	cb, 1st meth.	58.22	68.50	64.07	91.14	88.26	93.68	55.49	61.16	71.38	113.65	88.74	93.76
	cb, 2nd meth.	53.20	55.60	62.15	84.28	89.58	89.85	55.57	71.99	66.58	166.99	173.84	163.57
	pls, 1st meth.	65.12	60.97	67.35	88.51	99.98	92.66						
	pls, 2nd meth.	53.85	55.95	60.81	84.57	88.77	89.46						

it is possible to change the sign of the estimated coefficient of scale, and apply it to scale structural coefficients or latent scores for SEM indeterminacy. An alternative, and in our view preferable, approach is to consider the other estimator of the MIMIC or another calibration method to predict the IE.

The last robustness check examines the source of poorer predictive performances of the MIMIC approach when model B is calibrated by the second method. We find that the bias is due to the omission of a relevant variable (i.e. x_6) in the calibration Eq. 10.[45] Accordingly, if we include all the relevant variables (i.e. Eq. 10 = Eq. 2). The predictions of Model B based on the second calibration method become reliable ($R^2 \geq 0.98$ in 95% with $T=25$, see Online Appendix G) and as well as the first calibration approach.

In conclusion, if there are many available exogenous values as the number of relevant variables correlated with the reference indicator plus one, then the second

---

[45] We omitted this variable to preserve comparability in terms of specification and exogenous values in the assessment of the three calibration methods (i.e. we apply the second method using the same regression with $e=2$ and $e=3$).

Table 4 Goodness of fit of the predicted IE (R^2) (T = 25)

		Est	Calibration	Dataset composition	N (coverage)	5th percentile	10th percentile	Median	Mean	Standard Deviation
Exogenous Values of IE = 2	Model A	CB-SEM	1st meth.	Full Sample	3675 (0.613)	0.605	0.903	0.995	-7.588	318.3
				Min-Max	3487 (0.581)	0.865	0.949	0.995	0.961	0.216
			2nd meth.	Full Sample	3675 (0.613)	0.977	0.990	0.998	0.257	34.39
				Min-Max	3637 (0.606)	0.982	0.991	0.998	0.989	0.069
		PLS-SEM	1st meth.	Full Sample	6000 (1)	0.236	0.810	0.99	-29.998	2056
				Min-Max	5529 (0.922)	0.822	0.922	0.992	0.955	0.169
			2nd meth.	Full Sample	6000 (1)	0.956	0.974	0.993	0.53	19.10
				Min-Max	5919 (0.987)	0.963	0.976	0.993	0.984	0.060
	Model B	CB-SEM	1st meth.	Full Sample	3135 (0.523)	0.577	0.896	0.995	-101.7	5391
				Min-Max	2971 (0.495)	0.863	0.949	0.995	0.96	0.225
			2nd meth.	Full Sample	3135 (0.523)	-2.109	0.121	0.975	-233.1	7279
				Min-Max	2641 (0.440)	0.547	0.770	0.984	0.908	0.246
Exogenous Values of IE = 3	Model A	CB-SEM	1st meth.	Full Sample	3675 (0.613)	0.977	0.988	0.997	0.976	0.589
				Min-Max	3656 (0.609)	0.978	0.988	0.997	0.991	0.083
			2nd meth.	Full Sample	3675 (0.613)	0.993	0.994	0.998	0.934	3.525
				Min-Max	3671 (0.612)	0.993	0.994	0.998	0.997	0.005
		PLS-SEM	1st meth.	Full Sample	6000 (1)	0.953	0.976	0.994	0.978	0.202
				Min-Max	5935 (0.989)	0.957	0.977	0.994	0.986	0.044
			2nd meth.	Full Sample	6000 (1)	0.978	0.983	0.994	0.938	3.810
				Min-Max	5989 (0.998)	0.978	0.983	0.994	0.992	0.008
	Model B	CB-SEM	1st meth.	Full Sample	3135 (0.523)	0.977	0.988	0.997	0.971	0.614
				Min-Max	3118 (0.520)	0.979	0.988	0.997	0.991	0.090
			2nd meth.	Full Sample	3135 (0.523)	0.612	0.898	0.994	-0.304	20.81
				Min-Max	2833 (0.472)	0.934	0.964	0.995	0.980	0.085

Grey cells indicate that at least 95% of estimated models have $R^2>0.95$. n indicates the number of estimated models. these values are equal to the total estimated models for ols and pls, while for cb estimator n is equal to the number of models in which the ml estimator converges

calibration method is expected to have adequate performances also in more complex MIMIC specifications as Model B.

4 An empirical application: the Italian underground economy

In this section, we apply the MIMIC approach to predict the Italian underground[46] economy (UE). We choose the Italian UE as a real benchmark to apply the MIMIC because the ISTAT is one of the few national institutes of statistics that makes publicly available detailed estimates of the NOE and with a wide sample size (form 2011 to 2018). This allows comparing predictions based on different values of parameter *e* (i.e. exogenous estimates of the UE) as analyzed in the MC simulation. It is outside the scope of this article to investigate the causes and characteristics of Italian UE therefore, in the following, we only focus on the application of the MIMIC approach.

---

[46] In this section, we utilize the adjective "underground" instead of "informal" because we use as reference indicator the estimates of "underground value added" calculated by the Italian National Institute of Statistics (ISTAT) for measuring NOE.

4.1 MIMIC model for the Italian underground economy

This section specifies six (nested) MIMIC models to estimate the Italian UE over the period 1995–2020. With the exclusion of Model V that can be estimated only by CB because of the inclusion of a control variable in the measurement equation of the reference indicator (as Model B in MC analysis), we estimate the MIMIC by both SEM estimators.

Following the empirical literature on the causes of IE, we specify four several structural models of the MIMIC and two measurement models.[47] The structural models include: three proxies of the tax burden, i.e. direct (x_{1a}), indirect (x_{1b}) or total (x_1) as a percentage of VA; an index of the flexibility of labor market calculated as the ratio between wage and salary workers whose jobs have a pre-determined termination date and all wage and salary workers in working age (x_2); an index of self-employment calculated as a ratio between full-time equivalent self-employed and total full-time equivalent employment (x_3); an index of labor cost measured as a ratio between employers' social contributions and domestic compensation of employees (x_4); Given that fluctuations in investments are a predictor of upcoming business activity, we use the ratio between the gross fixed capital formation and VA as a proxy of the effect of the future business cycle on informality (x_5); As a proxy of regulatory burden and the presence of the public sector in the market, we consider the ratio between values added of public administration, defense and compulsory social security and total VA (x_6).

As the measurement models concern, we consider four potential indicators of the UE: the "reference indicator" is a proxy of the UE based on the NA approach, i.e. ISTAT's (2019) estimates of the UE for the period 2011–2018 (y_1); ISTAT's estimates of undeclared work (y_2); official estimates of tax-gap (y_3); a proxy of informal employment in accommodation and food service activities (y_4). In the first (y_1) and second (y_2) measurement equation of MIMIC V, we also include the total amount of worked hours (x_7) to control for the effect of the business cycle.

4.2 Estimates and calibration of the MIMIC model

We estimate six nested MIMIC models applying both SEM-estimators and the three calibration methods. We report estimates when only two exogenous values of the UE are available (i.e. $t_1^* = 2017$–2018).[48] Given that in CB-SEM the likelihood-ratio test is derived under the assumption that observed variables are normally distributed, we preliminary test for multivariate normality. According to Henze-Zirkler and Doornik–Hansen tests, we reject the hypothesis of multivariate normality, therefore

---

[47] See Online Appendix H, for definitions, descriptive statistics and sources of data. All the variables are included as ratio and in percentage points, with exclusion of x_7. The variable measured in monetary values are divided by Gross Value Added (VA) to be consistent with the NA approach that primarily estimates the underground VA.

[48] For the sake of brevity, we do not report the analysis also with $t_2^* = 2014$–2018; $t_3^* = 2011$-2018. The results are robust to these changes

we apply Satorra–Bentler (SB) estimator to adjust standard errors and the chi-squared test to make them robust to nonnormality.

For an overall assessment of the CB-MIMICs, the SB chi-squared test points out as only model VI the hypothesis of a perfect fit between model-implied covariance matrix and sample variance matrix cannot be rejected at 0.05 level of statistical significance. As the assessment of the PLS models concerns, this estimator lacks an index for overall model fit, therefore selecting the best specification among alternative model structures may be problematic. However, we report three goodness-of-fit indexes that measure, separately, the reliability of structural and measurement models of the latent variable UE = η_1.[49]

These statistics indicate that all the MIMICs provide a good representation of the latent construct and (in terms of statistical significance) the coefficients of MIMIC III, IV and VI are fairly robust to SEM estimators.[50] In relative terms, model VI has the best statistical performances. Table 5 summarizes these results. The last step of the MIMIC approach consists in calibrating the coefficients. The estimates of $^{ols\_est}\widehat{\rho}_1$ (method 1) and $^{ols}\widehat{\lambda}_1^*$ (method 2) are used to "scale" the coefficients of Table 5, while $^{ols\_est}\widehat{\phi}_1$ (method 3) is an estimate of the standard deviation of the UE.[51]

In this analysis, we encounter the problem of an inverted trend in predicted IE based on Model III, IV and VI when the second calibration method is applied to t_2^* or t_3^*. Following previous guidelines, we detect the inverted trend in four ways. First, we include in Eq. 10 the omitted relevant variables. Second, identifying the "right" trend by applying the third calibration method (i.e. looking at the sign of the estimated coefficient $^{ols\_est}\widehat{\phi}_1$ in Eq. 16). Third, given that we expect a positive correlation between the reference indicator (i.e. the index of UE based on ISTAT estimates) and the latent variable, then the negative sign of the coefficient of scale (i.e. $^{ols}\widehat{\lambda}_1^* < 0$) signals an inverted trend. Fourth, by comparing the plots of predicted UE and the exogenous values. All these checks make clear that the MIMIC predictions based on models with $^{ols}\widehat{\lambda}_1^* < 0$ have a reversed trend compared to the true values of exogenous estimates.[52] In conclusion, according to these concerns, multiplying $^{ols}\widehat{\lambda}_1^*$ by "−1", and use "$-^{ols}\widehat{\lambda}_1^*$" to correct the signs of structural coefficients allows to solve the issue of inverted trend. Figure 3 shows the estimates of the Italian UE as a percentage of VA by models III, IV, V, VI applying both SEM estimators, the three

---

[49] Following the literature on PLS evaluation (e.g. Hair et al., 2019) we report, for the structural model, the R^2-adjusted, for measurement model, the Average Variance Extracted (AVE)—where a value larger than 0.5 is considered as a signal of good convergent validity of reflective latent construct—and the index of internal consistency reliability (ρ_A)—where a value lies between 0.70 and 0.95 indicates that the reflective measurement model provides a good representation of latent construct -.

[50] The coefficients differ in their absolute values because the PLS-estimator employs standardized variables while the CB algorithm uses raw data. To convert PLS coefficients to the same scale of CB-SEM, we have to multiply the i-th structural coefficient by $\sigma_{y_1}/\sigma_{x_i}$ and the j-th measurement coefficient by $\sigma_{y_1}/\sigma_{y_j}$, where σ_{y_1} is the standard deviation of the reference indicator.

[51] Appendix I reports the OLS estimates of the three calibration equations (Eqs. 6, 10, 16) for models III, IV V and VI based on two, five and eight exogenous values of the UE (i.e. $e = 2, 5, 8$).

[52] Tables I.2 and I.3 in Online Appendix I show that the reverse trend is due to the "apparent" negative sign of the coefficient of scale that, in fact, is not statistically significant.

periods of calibrations (highlighted by two vertical lines) and the two calibration approaches.

Figure 3 shows that the predicted values of the IE are robust to calibration methods, model specifications and SEM estimators (except for models IV, V and VI estimated by CB algorithm with the first calibration method over the period t_1^* =2017–18).

In line with the previous general conclusions of the MC analysis, the second calibration method offers better chances to correctly estimate the UE and negligible differences emerge comparing predictions from CB to PLS algorithms. We do not encounter problems of convergence in CB but, when we apply the second method of calibration over the periods t_2^* or t_3^*, predicted values have a reversed trend. Although this issue is solved by changing the sign of the coefficient of scale (e.g. Figure 3 shows predicted values after this adjustment), for caution, we suggest to select predicted values based on another calibration approach (e.g. method 1 or 3). In conclusion, Fig. 4 shows the estimates of the Italian UE as a percentage of GDP and in billions of euros as predicted by the MIMIC VI with both the calibration methods and compares them to the ISTAT (i.e. NA) estimates of the UE.

5 Conclusions

The exercise to estimate something, whose nature is not observable, is very complicated, and seen with skepticism from some quarters of the statistical accountants and applied economists. Some scholars (e.g. Breusch, 2016; Feige, 2016; Kirchgässner, 2016; Slemrod & Weber, 2012) and some institutions (e.g. OECD, 2002) assess one of the most employed econometric methods to estimate IE, the MIMIC approach, as unreliable. Dybka et al. (2019) group these criticisms into three main areas: (a) the way in which the MIMIC framework has been applied, including non-compliance with the academic standards of transparency in exposition, replicability or conservatism in formulating conclusions; (b) the idea of applying the MIMIC approach to the IE measurement; (c) the issues related to specification, identification and estimation of the particular model used to estimate the IE. Aware of these criticisms, we try to describe (as transparently as possible) the estimation strategy and propose practical suggestions on how to address the main difficulties of the MIMIC approach. We aim to contribute to this literature from a twofold perspective.

From a theoretical viewpoint, we point out as, by considering the MIMIC model as one of the possible specifications of the SEM, then the criticisms of the inappropriateness of the MIMIC to estimate the IE can be downgraded from a general criticism towards the approach to a critique of a particular model specification used in that empirical application due to data availability and researcher's hypotheses. We have shown how it is possible to implement more complex and complete (economic and measurement) theories of IE into a SEM model (e.g. Model A vs. Model B).

From an empirical viewpoint, differently from other fields of applied economics, it is not possible to test the goodness-of-fit of predicted values because full scope information for the (estimated) IE is never available. In this article, we attempt to

overcome this difficulty, by simulating tens of thousands of IE, in order to compare SEM predictions to "actual" values. In particular, the MC analysis has been set to simulate "realistic" values of observed variables and IE to increase the external validity of results. A second empirical contribution is applying, for the first time in this area of research, the PLS algorithm to estimate a MIMIC model and comparing PLS predictive performances to standard CB estimator. A third empirical contribution consists in analyzing, on the one hand, the statistical performances (in terms of bias, consistency and efficiency) of the SEM estimators and, on the other hand, the goodness-of-fit of the predicted IE. Keeping these two performances distinct is relevant because while the estimator properties depend on the suitability of the SEM approach to analyze the IE, the predictive performances mainly deal with the reliability of calibration approaches. On this issue, we have proposed three (complementary) approaches that aim to address the most important shortcoming of the MIMIC approach, i.e. how to convert the index of the latent variable to the actual measure of the IE.

The main findings of this research are that CB (normally) dominates PLS.[53] The main merit of the PLS estimator is that, differently from the ML approach which often encounters problems of non-convergence, PLS (practically) always converges and the cost in terms of lower accuracy of the predictions is acceptable.[54] The preference for CB is also motivated by the fact that the PLS does not have an adequate global goodness-of model fit measure (such as chi-square for CB-SEM) therefore the scholar does not have a clear criterion on which is the best model specification. However, according to our simulations and empirical application on the Italian UE, the differences between estimators are rather small.

As calibration methods concern, the second method (based on measurement model) has generally better performances than the first method (based on the structural equation). The MC analysis has pointed out as, by increasing the number of exogenous estimates, there is a significant improvement in accuracy and consistency of estimations.

In conclusion, a clear answer on how much we can trust in the SEM or MIMIC estimates of the IE is not easy. As Schneider (1997) stated, each approach has its strengths and weaknesses and can provide specific insights and results. We agree with this statement. In particular, we believe that the main advantage of the MIMIC—i.e. its flexible framework that allows defining causes and indicators of the IE in line with the features of the analyzed economy and availability of data—may

---

[53] For instance, with reference to the "classical" MIMIC specification (model A), we find that $R^2 > 0.95$ in more than 95% of the predicted IE when it is calibrated with two exogenous values and, this statistics further increases if $e = 3$ ($R^2 > 0.99$).

[54] With reference to the possibility to encounter indefinite matrix problems, Bollen and Long (1993) points out as some Monte Carlo studies demonstrate that their frequency increases when the data provides relatively little information (e.g. small sample size, few observed indicator variables, small factor loadings, missing values). In our simulations, that do not have missing values, the issue of non-convergence of the CB algorithm has occurred in about 60% of simulations. Therefore, we guess that, for empirical analysis where data availability is a relevant issue, applying the PLS-SEM, may be the only option to predict the IE by the MIMIC.

Table 5 Estimated Coefficients of the CB-SEM and PLS-SEM Models (2 Exog. Values, $t_1^* = 2017\text{–}2018$)

Models →		I CB	I PLS	II CB	II PLS	III CB	III PLS	IV CB	IV PLS	V CB	VI CB	VI PLS
Potential Causes												
Direct Tax B. (x_{1a})	γ_{1a}	-0.02	0.005	--	--	--	--	--	--	--	--	--
Indirect Tax B. (x_{1b})	γ_{1b}	-0.021	0.074	--	--	--	--	--	--	--	--	--
Total Tax B. (x_1)	γ_1	--	--	-0.025**	0.025	--	--	--	--	--	--	--
Flex. Lab.(x_2)	γ_2	0.01	-0.093	0.01	-0.062	--	--	--	--	--	--	--
Self-Empl.(x_3)	γ_3	0.104***	0.487	0.106***	0.535	0.091***	0.586***	0.081***	0.584***	0.078***	0.080***	0.655***
Lab. Cost (x_4)	γ_4	0.052**	0.338	0.052**	0.284	0.066***	0.284**	0.068***	0.285**	0.088***	0.069***	0.214**
Investments (x_5)	γ_5	-0.003	0.007	-0.005	0.004	--	--	0.016*	0.004	--	0.016*	0.07
Size P.A. (x_6)	γ_6	-0.237***	-0.368***	-0.228***	-0.371***	-0.238***	-0.357***	-0.216***	-0.356***	-0.257***	-0.213***	-0.306***
Measurement (Outer) Models												
Istat_Under. VA (y_1)		--		--		--		--		--	--	
UE=η_1	λ_1	1(constr.)	0.967***	1(constr.)	0.967***	1(constr.)	0.968***	1(constr.)	0.968***	1(constr.)	1(constr.)	0.975***
Off.W.hours (x_7)	γ_7	--	--	--	--	--	--	--	--	0.033***	--	--
Constant		9.871***	--	9.980***	--	9.354***	--	9.115***	--	7.828***	9.073***	--
Undeclared Work (y_2)		--		--		--		--		--	--	
UE=η_1	λ_2	1.62	0.453	1.642	0.45	1.804	0.447	6.240***	0.447	3.091***	--	--
Self-Empl.(x_3)	γ_8	--	--	--	--	--	--	-0.690***	0.23	-0.437***	--	--
Investments(x_5)	γ_9	--	--	--	--	--	--	-0.292***	-0.441***	--	--	--
Off.W.hours (x_7)	γ_{10}	--	--	--	--	--	--	--	0.344***	-0.304***	--	--
Constant		11.27***	--	11.39***	--	9.862**	--	22.54***	--	32.32***	--	--
Tax Gap %VA (y_3)		--		--		--		--		--	--	
UE=η_1	λ_3	6.280***	0.946***	6.299***	0.946***	6.383***	0.946***	7.022***	0.946***	5.884***	6.991***	0.925***
Investments(x_5)	γ_{11}	--	--	--	--	--	--	-0.184***	0.249	--	-0.182***	0.249
Constant		-10.117	--	-9.477*	--	-13.68***	--	-13.35***	--	-12.65***	-13.58***	--
Und.Work Acc.Food(y_4)		--		--		--		--		--	--	
UE=η_1	λ_4	17.34***	0.784***	17.26***	0.785***	17.548**	0.786***	24.91***	0.787***	18.413***	23.892***	0.834***
Self-Empl..(x_3)	γ_{12}	--	--	--	--	--	--	2.018***	0.996***	2.375***	2.129***	0.996***
Lab. Cost (x_4)	γ_{13}	--	--	--	--	--	--	-3.151***	-0.278	-2.955***	-3.101***	-0.278
Investments(x_5)	γ_{14}	--	--	--	--	--	--	0.637***	--	--	--	--
Constant		32.925	--	35.001*	--	23.323	--	19.421	--	13.429	16.938	--
M. Assessment												
Satorra-Bentler χ^2 test (CB)		134.993	--	129.353	--	68.321	--	21.218	--	46.89	8.778	--
Satorra-Ben. χ^2 p-value (CB)		0.000	--	0.000	--	0.000	--	0.017	--	0.001	0.344◊	--
Degrees of freedom◊◊		23	--	23	--	12	--	12	--	19	8	--
R²-adjusted (UE=η_1) (PLS)		--	0.892⁺	--	0.897⁺	--	0.487	--	0.907⁺	--	--	0.921⁺
AVE (UE=η_1) (PLS)		--	0.663⁺	--	0.662⁺	--	0.662⁺	--	0.662⁺	--	--	0.834⁺
ρ_Λ(UE=η_1) (PLS)		--	0.903⁺	--	0.906⁺	--	0.908⁺	--	0.908⁺	--	--	0.910⁺

***, **, * represents *p* values lower than 1%, 5% and 10%; $\lozenge$ means that we fail to reject at 5% level, the null hypothesis of the Satorra–Bentler scaled chi-squared test, indicating good fit in CB-SEM; $\lozenge\lozenge$ The degrees of freedom are determined by $0.5(p+q)(p+q+1) + m - t$, where "*p*" is the number of indicators, "*q*" the number of causes, "*m*" the number of means and intercepts and "*t*" is the number of free parameters; ⁺ Means good fitting in PLS-SEM

be also its main source of vulnerability. Indeed, the theoretical construct, that the scholar labels as "informal economy" could have other potential definitions. This objection, pointed out by Helberger and Knepel (1988) Giles and Tedds (2002), Dell'Anno (2003), remains difficult to overcome in general terms, as it due to the theoretical assumptions behind the choice of variables and empirical limitation on the availability of data and should be evaluated on a case-by-case basis.

As stated in the introduction section, the MIMIC method is not an exception to the rule that no statistical method turns bad data into good data. On the contrary, we consider the quality of exogenous estimates of the informality as crucial to get reliable estimates of the IE using this approach. The rationale is that external estimates are employed to calibrate the model (i.e. estimate mean and variance of the predicted IE). Accordingly, we suggest calibrating the MIMIC by using exogenous values obtained by the NA approach rather than econometric methods to prevent two problems. First, applying an indirect (e.g. MIMIC, monetary, physical input) approach may generate a chain of untraceable references, making the identification of the primary source demanding and, consequently, reducing transparency and replicability. Second, layering econometric approaches means that any impreciseness (related to calibration, estimation, etc.) can multiply in time leading to vast distortions of the IE estimates. However, this suggestion does not significantly reduce the applicability of the MIMIC approach to estimate the IE because, the three calibration approaches
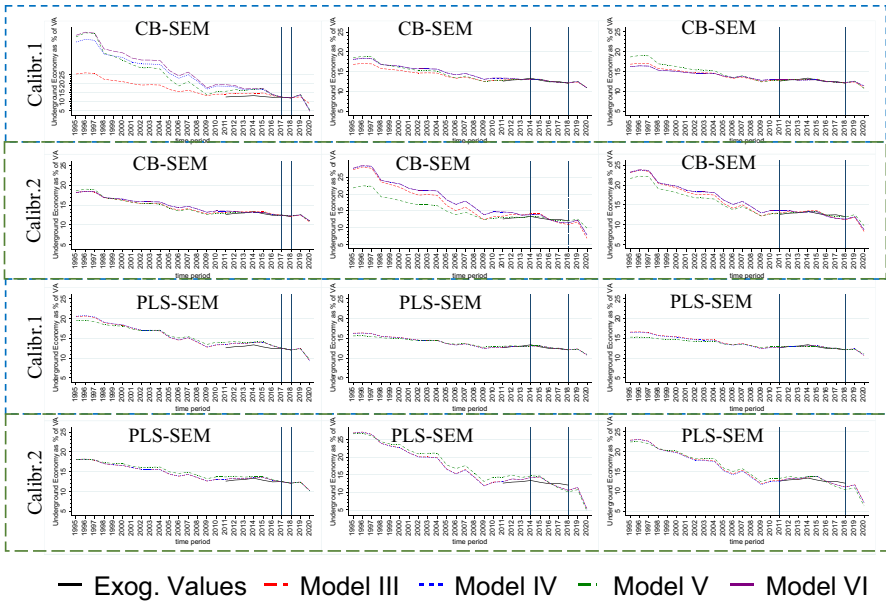
Fig. 3 Plots of MIMIC III, IV, V and VI (t_1^*=2017–18; t_2^*=2014–18; t_3^*=2011–18)
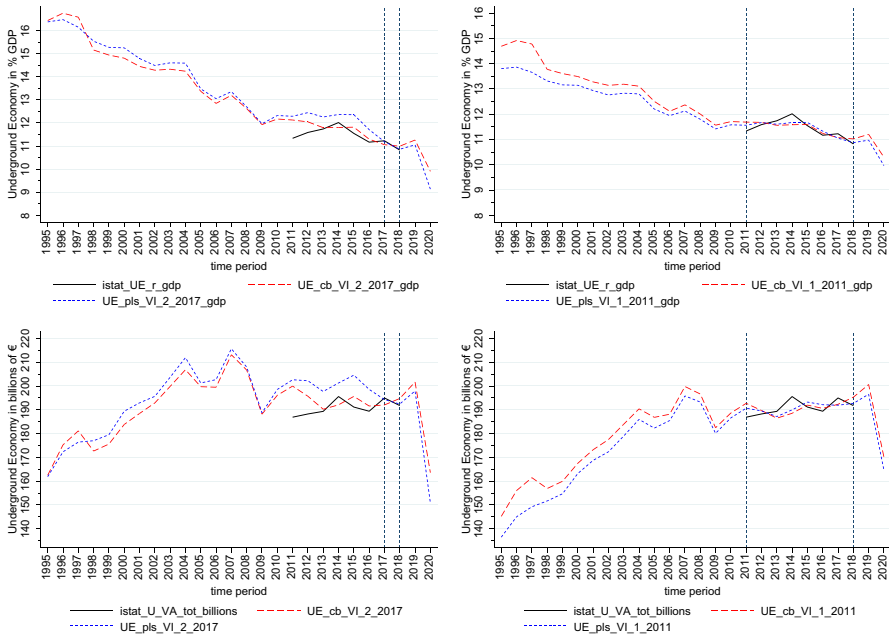

Fig. 4 Italian UE Model VI (2nd method, t^* = 2017–18; 1st method, t^* = 2011–18)

proposed in this paper require at least two values (or preferably more), which in fact is not challenging. Although many statistical offices do not frequently publish their estimates of the NOE, they sometimes publish occasional reports with estimates for selected years. In such case, an accurately calibrated MIMIC model can be used to extend the range of NA estimates.

In response to the skepticism of some scholars, we hope that more and more scholars will accept the amazing challenge of measuring the unmeasurable economy by looking into advantages and shortcomings of the SEM approach. We do not consider the SEM approach as the most reliable method to estimate the UE[55] nor that all the difficulties are solved (e.g. how to deal with cointegrated variables, how to calibrate panel data, how adequately address the issue of endogeneity, etc.). In spite of this, we believe that the flexibility of its framework and the inclusiveness to the other econometric approaches[56] make the SEM the most promising econometric approach to fill the gap between the vast scholar's demand of quantitative knowledge on the IE and the scarce availability of NA estimates.

Declarations

Conflict of interest The author has no relevant financial or non-financial interests that are directly or indirectly related to this article to disclose.

References

Andrews, I., Gentzkow, M., & Shapiro, J. (2020). Transparency in structural research. *Journal of Business & Economic Statistics, 38*(4), 711–722.

---

[55] It is the national accounting approach to estimate the NOE when applied by the national institutes of statistics because they have better data and more resources.

[56] E.g. we can use the estimates of the IE based on other indirect and/or direct methods as observable indicators, the estimates of the NOE based on NA as exogenous values for calibration.

Biles, J. J. (2009). Informal work in Latin America: Competing perspectives and recent debates. *Geography Compass, 3*(1), 214–236.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Sage.

Breusch, T. (2016). Estimating the underground economy using MIMIC models. *Journal of Tax Administration, 2*(1), 1–29.

Buehn, A., & Schneider, F. (2012). Shadow economies around the world: Novel insights, accepted knowledge, and new estimates. *International Tax and Public Finance, 19*(1), 139–171.

Cassar, A. (2001) An index of the underground economy in Malta. *Bank of Valletta Review, 23, Spring, 44–62.*

de Soto, H. (1989). *The other path*. Harper and Row.

Dell'Anno, R. (2003) Estimating the shadow economy in Italy: A structural equation approach», *University of Aarhus, Department of Economics, DK. Working paper 2003–7.*

Dell'Anno, R. (2016). Analyzing the determinants of the shadow economy with a 'separate approach'. An Application of the relationship between inequality and the shadow economy. *World Development, 84*, 342–356.

Dell'Anno, R. (2021). Theories and definitions of the informal economy: A survey. *Journal of Economic Surveys*. https://doi.org/10.1111/joes.12487

Dell'Anno, R., Schneider, F. (2006) *Estimating the underground economy: a response to T. Breusch's critique*. Discussion Paper, 2006–02, Department of Economics, Johannes Kepler University of Linz, Austria.

Dell'Anno, R., Schneider, F. (2009) A complex approach to estimate shadow economy: The structural equation modelling. In Faggini M., Lux T. (eds) *Coping with the complexity of economics. Springer-Verlag New Economic Windows series*, pp.111–130.

Dell'Anno, R., & Davidescu, A. (2019). Estimating shadow economy and tax evasion in Romania. A comparison by different estimation approaches. *Economic Analysis and Policy, 63*, 130–149.

Dybka, P., Kowalczuk, M., Olesiński, B., Torój, A., & Rozkrut, M. (2019). Currency demand and MIMIC models: Towards a structured hybrid method of measuring the shadow economy. *International Tax Public Finance, 26*, 4–40.

Feige, E. L. (2016). Reflections on the meaning and measurement of unobserved economies: What do we really know about the "shadow economy"? *Journal of Tax Administration, 2*(1), 93–107.

Frey, B., & Weck-Hanneman, H. (1984). The hidden economy as an "Unobservable" variable. *European Economic Review, 26*(1), 33–53.

Giles, D.E.A., Tedds, L.M. (2002). Taxes and the Canadian underground economy. *Canadian Tax paper n.106. Canadian Tax Foundation*. Toronto.

Goodhue, D. L., Lewis, W., & Thompson, R. (2012a). Comparing PLS to regression and LISREL: A response to Marcoulides, Chin, and Saunders. *MIS Quarterly, 36*(3), 703–716.

Goodhue, D. L., Lewis, W., & Thompson, R. (2012b). Does RLS have advantages for small sample size or non-normal data? *MIS Quarterly, 36*(3), 981–1001.

Gould, W. W., Pitblado, J. S., & Poi, B. P. (2010). *Maximum likelihood estimation with Stata* (4th ed.). Stata Press.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24.

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.

Quiros-Romero G., T.F. Alexander, Ribarsky J. (2021). *Measuring the Informal Economy IMF Policy Paper No. 2021/002.*

ISTAT (2019). *Non observed economy in national accounts*. Press release. Rome, Italy. (from https://www.istat.it/it/files//2019/10/Non-observed-economy-in-national-accounts-2017-EN.pdf).

Jöreskog, K., & 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*, 631–639.

Kirchgässner, G. (2016). On estimating the size of the shadow economy. *German Economic Review, 18*(1), 99–11.

Medina, L., & Schneider, F. (2020). Shedding light on the shadow economy. *World Economics Journal, 21*(2), 25–82.

OECD (2002). *Handbook for measurement of the Non Observed Economy*, Paris.

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research, 69*(10), 3998–4010.

Schneider, F. (1997). The shadow economies of Western Europe. *Economic Affairs, 17*(3), 42–48.

Schneider, F., & Enste, D. H. (2000). Shadow Economies: Size, causes, and consequences. *Journal of Economic Literature, 38*(1), 77–114.

Slemrod, J., & Weber, C. (2012). Evidence of the invisible: Toward a credibility revolution in the empirical analysis of tax evasion and the informal economy. *International Tax and Public Finance, 19*, 25–53.

Thomas, J. (1992). *Informal economic activity*. Handbooks in Economics, London, Harvester Wheatsheaf.

Ulyssea, G. (2020). Informality: Causes and consequences for development. *Annual Review of Economics, 12*(1), 525–546.

Zhou, F., & Oostendorp, R. (2014). Measuring true sales and underreporting with matched firm-level survey and tax office data. *The Review of Economics and Statistics, 96*(3), 563–657.